

### Εθνικό Μετσοβίο Πολυτεχνείο

Σχολή Ηλεκτρολογών Μηχανικών και Μηχανικών Υπολογιστών Τομέας Συστηματών Μεταδοσής Πληροφορίας και Τεχνολογίας Υλικών

### Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

### ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Ιωάννης Γ. Ζώρζος

Αθήνα, Μάιος 2025

Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

### ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ



#### ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ & ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΣΥΣΤΗΜΑΤΩΝ ΜΕΤΑΔΟΣΗΣ ΠΛΗΡΟΦΟΡΙΑΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ ΥΛΙΚΩΝ

### Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

### ΙΩΑΝΝΗΣ ΖΩΡΖΟΣ

Εγκρίθηκε από την επταμελή επιτροπή στις 16 Μαΐου 2025.

Γ. Ματσόπουλος Καθηγητής Ε.Μ.Π.

Δ.Δ. Κουτσούρης Ομ. Καθηγητής Ε.Μ.Π.

Π. Ασβεστάς Καθηγητής, ΠΑ.Δ.Α

Π. Τσανάκας Καθηγητής Ε.Μ.Π.

Ε. Χριστοφόρου Καθηγητής Ε.Μ.Π.

Α. ΠαναγόπουλοςΚαθηγητής Ε.Μ.Π.

Ι. Καλαντζής Καθηγητής, ΠΑ.Δ.Α

Αθήνα, Μάιος 2025

Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

### Ιωάννης Ζώρζος

Διδάκτωρ Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Ιωάννης Ζώρζος, 2025 Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της,εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου. Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

Στην πολυαγαπημένη μου σύζυγο, Ζωή

# Περίληψη

Το ηλεκτροεγκεφαλογράφημα (ΗΕΓ) αποτελεί μια νευροφυσιολογική μέθοδο με υψηλή χρονική ανάλυση, κρίσιμη για την κατανόηση της εγκεφαλικής λειτουργίας και την ανάπτυξη διεπαφών εγκεφάλου-υπολογιστή (BCI). Παρά την πλούσια πληροφορία που περιέχουν τα δεδομένα ΗΕΓ, η ανάλυση και ερμηνεία τους παρουσιάζουν σημαντικές προκλήσεις, ιδίως όσον αφορά την εξαγωγή ακριβούς χωρικής πληροφορίας από τους αισθητήρες του κρανίου και την κατανόηση των υποκείμενων νευρωνικών μηχανισμών που οδηγούν σε συγκεκριμένες αποκρίσεις ή συμπεριφορές. Ταυτόχρονα, η αυξανόμενη χρήση μοντέλων βαθιάς μάθησης (DL) για την ταξινόμηση σημάτων ΗΕΓ, ενώ οδηγεί σε εντυπωσιακά επίπεδα απόδοσης, συχνά δημιουργεί "μαύρα κουτιά" που καθιστούν δύσκολη την αιτιολόγηση των προβλέψεων, περιορίζοντας την εμπιστοσύνη και την κλινική τους υιοθέτηση.

Η παρούσα διατριβή επιδιώκει να γεφυρώσει το χάσμα μεταξύ της υψηλής απόδοσης των μοντέλων Τεχνητής Νοημοσύνης (TN) σε δεδομένα ΗΕΓ και της ανάγκης για φυσιολογικώς έγκυρη και ανατομικά πληροφορημένη ερμηνευσιμότητα. Για τον σκοπό αυτό, αναπτύχθηκε ένα ολοκληρωμένο πλαίσιο που συνδυάζει προηγμένες τεχνικές ανάλυσης σήματος, εντοπισμού πηγών και εξηγήσιμης βαθιάς μάθησης. Το έργο διαρθρώθηκε σε τρεις κύριους άζονες. Αρχικά, διερευνήθηκε η αξία της ανάλυσης χρόνου-συχνότητας (Time-Frequency Analysis) ως ισχυρού εργαλείου εξαγωγής χαρακτηριστικών, επιδεικνύοντας την αποτελεσματικότητά της σε συνδυασμό με ένα υπολογιστικά αποδοτικό συνελικτικό νευρωνικό δίκτυο για την ταξινόμηση σημάτων ΗΕΓ σε ένα πλαίσιο ανίχνευσης νοητικής κόπωσης. Αυτό το τμήμα ανέδειξε τη δυνατότητα επίτευξης υψηλής ακρίβειας (97%) με χαμηλό υπολογιστικό κόστος, εντοπίζοντας παράλληλα, μέσω τεχνικών ερμηνευσιμότητας όπως το SHAP, φασματικά πρότυπα (στις ζώνες θήτα και άλφα) συμβατά με τη βιβλιογραφία της κόπωσης, υπογραμμίζοντας τη χρησιμότητα της ανάλυσης TFR και των εξηγήσιμων μοντέλων πέρα από την κεντρική εφαρμογή BCI.

Στη συνέχεια, παρουσιάστηκε μια κριτική ανασκόπηση των τρεχουσών μεθόδων εντοπισμού πηγών ΗΕΓ (EEG Source Localization). Αναλύθηκαν οι θεωρητικές βάσεις, οι διάφορες κατηγορίες αλγορίθμων (π.χ. beamformers, αναλύσεις ελάχιστης νόρμας), οι εφαρμογές και οι εγγενείς περιορισμοί τους. Η ανασκόπηση αυτή θεμελίωσε την αναγκαιότητα της μετάβασης από την ανάλυση σε επίπεδο αισθητήρων σε επίπεδο εγκεφαλικών πηγών για την επίτευξη ακριβούς ανατομικής ερμηνείας, δικαιολογώντας την επιλογή μιας προσέγγισης εντοπισμού πηγών (beamformer) για την εκτίμηση φλοιϊκής δραστηριότητας στην κύρια εφαρμογή της διατριβής.

Το τρίτο και κεντρικό τμήμα της εργασίας εφάρμοσε τον συνδυασμό των παραπάνω τεχνικών σε ένα σύστημα BCI νοερής κίνησης, ενσωματώνοντας τεχνικές ερμηνευσιμότητας (SHAP, Gradients). Χρησιμοποιώντας χαρακτηριστικά που Integrated εξήχθησαν από πηγαιοτοποθετημένα δεδομένα ΗΕΓ (με χρήση beamformer και παρσελοποίηση Desikan-Killiany) ως είσοδο σε ένα συνελικτικό νευρωνικό δίκτυο, επιτεύχθηκε ανταγωνιστική ακρίβεια στην ταξινόμηση νοερών κινήσεων χεριών, ποδιών και γλώσσας. Το σημαντικότερο εύρημα προέκυψε από την εφαρμογή των τεχνικών ερμηνευσιμότητας: οι σημαντικότερες περιοχές που αναδείχθηκαν για κάθε κλάση νοερής κίνησης έδειξαν ισχυρή συμφωνία με την εδραιωμένη νευροεπιστημονική γνώση και τη σωματοτοπική οργάνωση του κινητικού φλοιού, αναπαράγοντας ουσιαστικά έναν λειτουργικό "homunculus" και τα φαινόμενα πλευρίωσης που παρατηρούνται σε μελέτες fMRI. Αυτή η σύγκλιση επιβεβαίωσε ότι ο ταξινομητής βασίζεται σε γνήσια νευροανατομικά πρότυπα και όχι σε τυχαία χαρακτηριστικά, άροντας την ασάφεια της ανάλυσης σε επίπεδο αισθητήρων και παρέχοντας ανατομικά εξειδικευμένες εξηγήσεις.

Τα αποτελέσματα αυτά έχουν σημαντικές συνέπειες τόσο για τις κλινικές εφαρμογές BCI όσο και για τη νευροεπιστήμη. Ενισχύουν την εμπιστοσύνη στη χρήση μοντέλων DL σε κλινικά περιβάλλοντα, επιτρέπουν την παροχή ανατομικά πληροφορημένης νευροανάδρασης και προσφέρουν ένα εργαλείο για την παρακολούθηση νευροπλαστικών αλλαγών. Σε επιστημονικό επίπεδο, το πλαίσιο μπορεί να λειτουργήσει ως εργαλείο επαλήθευσης γνωστών προτύπων αλλά και ανακάλυψης λιγότερο κατανοητών όψεων της εγκεφαλικής λειτουργίας. Παρά τους περιορισμούς (πεπερασμένη ακρίβεια εντοπισμού πηγών, ευαισθησία σε βαθιές πηγές, μέγεθος συνόλου δεδομένων), η διατριβή αποδεικνύει ότι ένα συνδυαστικό πλαίσιο μπορεί να επιτύχει υψηλή απόδοση χωρίς να παραμένει "μαύρο κουτί". Οι μελλοντικές κατευθύνσεις περιλαμβάνουν τη βελτίωση της χωρικής/χρονικής ανάλυσης, την ενσωμάτωση πολυτροπικών δεδομένων, την εφαρμογή προηγμένων τεχνικών ερμηνευσιμότητας (όπως ανάλυση αλληλεπιδράσεων) και την εφαρμογή σε κλινικούς πληθυσμούς και άλλες εργασίες ΗΕΓ. Συνολικά, η εργασία θεμελιώνει μια νέα γενιά διεπαφών εγκεφάλου-υπολογιστή και νευροεπιστημονικών εργαλείων, όπου η διαφάνεια και η κατανόηση των υποκείμενων νευρωνικών μηχανισμών είναι κεντρικής σημασίας.

**Λέξεις Κλειδιά:** Ηλεκτροεγκεφαλογραφία (ΗΕΓ), Διεπαφές Εγκεφάλου-Υπολογιστή (BCI), Βαθιά Μάθηση (Deep Learning), Ερμηνευσιμότητα (Explainability), SHAP, Integrated Gradients, Εντοπισμό πηγών (Source Localization), Ανάλυση Χρόνου-Συχνότητας (Time-Frequency Analysis), Νοερή Κίνηση (Motor Imagery), Νευροεπιστήμη (Neuroscience).

## Abstract

Electroencephalography (EEG) is a neurophysiological technique offering high temporal resolution, crucial for understanding brain function and developing Brain-Computer Interfaces (BCIs). Despite the rich information contained in EEG signals, their analysis and interpretation present significant challenges, particularly in extracting precise spatial information from scalp sensors and understanding the underlying neural mechanisms driving specific responses or behaviors. Simultaneously, the increasing use of deep learning (DL) models for EEG signal classification, while yielding impressive performance levels, often creates "black boxes" that make it difficult to justify predictions, limiting trust and clinical adoption.

This thesis aims to bridge the gap between the high performance of Artificial Intelligence (AI) models on EEG data and the need for physiologically valid and anatomically informed interpretability. To this end, a comprehensive framework combining advanced signal analysis techniques, source localization, and explainable deep learning was developed. The work was structured across three main axes. Firstly, the value of Time-Frequency Analysis was explored as a powerful feature engineering tool, demonstrating its effectiveness in conjunction with a computationally efficient Convolutional Neural Network for classifying EEG signals in a mental fatigue detection context. This part highlighted the potential for achieving high accuracy (97%) with low computational cost, while also identifying, through explainability techniques like SHAP, spectral patterns (in theta and alpha bands) consistent with fatigue literature, underscoring the utility of TFR analysis and explainable models beyond the central BCI application.

Subsequently, a critical review of current EEG Source Localization methods was presented. Their theoretical foundations, various algorithm categories (e.g., beamformers, minimum norm solutions), applications, and inherent limitations were analyzed. This review established the necessity of moving from sensor-level analysis to brain source level to achieve accurate anatomical interpretation, justifying the choice of a source localization approach (beamformer) for estimating cortical activity in the main application of the thesis.

The third and central part of the work applied the combination of these techniques to a motor imagery BCI system, integrating explainability techniques (SHAP, Integrated Gradients). By using features extracted from source-localized EEG data (using a beamformer and Desikan–Killiany parcellation) as input to a convolutional neural network, competitive accuracy in classifying motor imagery of hands, feet, and tongue was achieved. The most significant finding emerged from applying the explainability techniques: the most important regions highlighted for each motor imagery class showed strong agreement with established neuroscientific knowledge and the somatotopic organization of the motor cortex, essentially reproducing a functional "homunculus" and laterality phenomena observed in fMRI studies. This convergence confirmed that the classifier relies on genuine neuroanatomical patterns rather than random features, resolving the ambiguity of sensor-level analysis and providing anatomically specific explanations.

These results have significant implications for both clinical BCI applications and neuroscience. They enhance trust in the use of DL models in clinical settings, enable the provision of anatomically informed neurofeedback, and offer a tool for monitoring neuroplastic changes. At a scientific level, the framework can serve as a tool for verifying known patterns and discovering less understood aspects of brain function. Despite the limitations (finite source localization accuracy, sensitivity to deep sources, dataset size), the

thesis demonstrates that a combined framework can achieve high performance without remaining a "black box." Future directions include improving spatial/temporal resolution, integrating multimodal data, applying advanced explainability techniques (such as interaction analysis), and application to clinical populations and other EEG tasks. Overall, the work lays the foundation for a new generation of brain-computer interfaces and neuroscientific tools where transparency and understanding of underlying neural mechanisms are central.

**Keywords:** Electroencephalography (EEG), Brain-Computer Interfaces (BCI), Deep Learning, Explainability, SHAP, Integrated Gradients, Source Localization, Time-Frequency Analysis, Motor Imagery, Neuroscience.

# Ευχαριστίες

Η εργασία που παρουσιάζεται σε αυτή τη Διατριβή πραγματοποιήθηκε στο εργαστήριο Βιοϊατρικής Τεχνολογίας του Εθνικού Μετσόβιου Πολυτεχνείου, υπό την επίβλεψη του Καθηγητή, Γεώργιου Κ. Ματσόπουλου. Χωρίς τη συμβολή του, τις συμβουλές του και τη καθοδήγησή του, η εργασία δεν θα είχε πραγματοποιηθεί.

Κατά τη διάρκεια εκπόνησης αυτής, αρκετοί άνθρωποι συνέβαλαν. Θα ήθελα συγκεκριμένα να ευχαριστήσω τους συνεπιβλέποντες καθηγητές, κ. Δημήτριο-Διονύσιο Κουτσούρη και κ. Παντελεήμονα Ασβεστά που με τη παρουσία τους και με τα ερευνητικά σχόλιά τους βελτίωσαν την εργασία αυτή.

Ιδιαίτερες ευχαριστίες οφείλω στον Δρ. Ιωάννη Κάκκο, ο οποίος λειτούργησε ως πολύτιμος μέντορας, συνεργάτης και φίλος. Η συνεχής του ενθάρρυνση, η υποστήριξή του στις καλές αλλά και στις δύσκολες στιγμές, και η ικανότητά του να διατηρεί το ηθικό μου υψηλό υπήρξαν καθοριστικής σημασίας.

Ευχαριστώ θερμά τους συναδέλφους υποψήφιους διδάκτορες, Σταύρο Θεοφάνη Μηλούλη, Αικατερίνη Καραμπάση, Γεώργιο Μπότη, Θεόδωρο Βαγενά και Κωνσταντίνο Ρακτιβάν. Η συνεργασία, η αλληλοϋποστήριξη και οι στιγμές που μοιραστήκαμε στο εργαστήριο έκαναν αυτή τη διαδρομή πιο ευχάριστη και δημιουργική. Ήταν χαρά μου να δουλεύω μαζί σας.

Ευχαριστώ από καρδιάς τους γονείς μου και την αδελφή μου για την αγάπη, την υπομονή και τη συνεχή τους στήριξη όλα αυτά τα χρόνια. Η πίστη τους σε εμένα ήταν πάντα πηγή δύναμης.

Πάνω απ' όλους όμως, θέλω να εκφράσω την απέραντη αγάπη και ευγνωμοσύνη μου στη σύζυγό μου, **Ζωή**. Χωρίς την αμέριστη κατανόηση, την υπομονή, την ενθάρρυνση και την αγάπη της, η ολοκλήρωση αυτής της διατριβής δεν θα ήταν δυνατή. Αυτή η δουλειά είναι αφιερωμένη σε εκείνη.

### Περιεχόμενα

| Περίληψη   | 1           |
|--|-------------|
| Abstract   | 3           |
| Ευχαριστίες  | 5           |
| Πίνακας Σχημάτων   | 9           |
| Λίστα Πινάκων  | 10          |
| Extended Summary: Techniques for Improving the Explainability of Artificial Intellig<br>Systems in Electroencephalography Data | ence<br>12  |
| Chapter 1: Introduction  | 12          |
| Chapter 2: Prerequisites and Modern Techniques   | 13          |
| Chapter 3: Explainability using Time-Frequency Analysis – Mental Fatigue Detec   | ction<br>14 |
| Chapter 4: Modern EEG Source Localization Techniques: A Review   | 15          |
| Chapter 5: Explainability using Source Localization Technique in a Brain-Comp<br>Interface (BCI)                               | outer<br>17 |
| Chapter 6: Discussion and Epilogue   | 19          |
| 1. Κεφάλαιο 1 - Εισαγωγή   | 21          |
| 2. Κεφάλαιο 2 – Προαπαιτούμενα και σύγχρονες τεχνικές  | 23          |
| 2.1 Ηλεκτροεγκεφαλογράφημα   | 23          |
| 2.1.1 Τύποι ηλεκτροδίων ΗΕΓ  | 23          |
| 2.1.2 Τοποθέτηση ηλεκτροδίων ΗΕΓ και μοντάζ  | 24          |
| 2.1.3 Ρυθμοί και ταλαντώσεις ΗΕΓ   | 25          |
| 2.1.4 Τεχνουργήματα και θόρυβος ΕΕG  | 27          |
| 2.1.5 Διόρθωση και προεπεξεργασία τεχνουργημάτων   | 27          |
| 2.1.6 Ανάλυση χρόνου-συχνότητας - Time-Frequency Analysis  | 30          |
| 2.1.7 Απεικόνιση ηλεκτρικής πηγής - EEG Source Localization  | 32          |
| 2.2. Μηχανική και Βαθιά Μάθηση   | 40          |
| 2.2.1 Ταξινόμηση και Μετρικές Αξιολόγησης  | 42          |
| 2.3 Εργαλεία Επεξηγησιμότητας (XAI) – Έμφαση στο SHAP και το Gradient Expla  | ainer<br>43 |
| 3. Κεφάλαιο 3 – Επεξηγησιμότητα με χρήση ανάλυσης χρόνου συχνότητας  | 47          |
| 3.1 Εισαγωγή   | 47          |
| 3.2 Ανάλυση χρόνου-συχνότητας για την ανίχνευση ψυχικής κόπωσης  | 49          |
| 3.2.1. Συμμετέχοντες   | 49          |
| 3.2.2. Πειραματικός σχεδιασμός   | 49          |
| 3.2.3. Λήψη δεδομένων και προεπεξεργασία   | 50          |
| 3.2.4. Εξαγωγή χαρακτηριστικών   | 50          |

| 3.2.5. Μοντελοποίηση τεχνητής νοημοσύνης  | 51         |
|---|------------|
| 3.3 Αποτελέσματα μοντέλου χρόνου-συχνότητας (ανίχνευση ψυχικής κόπωσης)   | 53         |
| 3.3.1 Επιδόσεις   | 53         |
| 3.3.2. Επεξηγησιμότητα  | 56         |
| 3.4 Συζήτηση  | 57         |
| 4. Κεφάλαιο 4 – Σύγχρονες τεχνικές εντοπισμού πηγής   | 61         |
| 4.1 Σύγχρονες μέθοδοι και προβλήματα  | 62         |
| 4.1.1. Λύση εμπρόσθιου προβλήματος  | 62         |
| 4.1.1 Αντίστροφο πρόβλημα   | 64         |
| 4.2 Εφαρμογές ESI-κλινικής και γνωστικής έρευνας  | 66         |
| 5. Κεφάλαιο 5 – Επεξηγησιμότητα με χρήση τεχνικής εντοπισμού πηγής  | 73         |
| 5.1 Εισαγωγή  | 73         |
| 5.2 Σχετικές Εργασίες   | 74         |
| 5.3 Μεθοδολογία   | 76         |
| 5.4 Αποτελέσματα επεξηγησημότητας μοντέλου εντοπισμού πηγής και τμηματοποίr<br>(εφαρμογή σε Brain Computer Interface) | ισης<br>80 |
| 5.1.1 Απεικόνιση νοερής κίνησης αριστερού χεριού (Left Hand Imagery)  | 81         |
| 5.2.2 Απεικόνιση νοερής κίνησης δεξιού χεριού (Right Hand Imagery)  | 83         |
| 5.2.3 Απεικόνιση νοερής κίνησης ποδιών (Foot Imagery)   | 85         |
| 5.2.4 Απεικόνιση νοερής κίνησης γλώσσας (Tongue Imagery)  | 86         |
| 5.5 Συζήτηση  | 88         |
| 6. Συζήτηση και Επίλογος  | 95         |
| 6.1 Συνοπτικά Ευρήματα και Συνεισφορές  | 95         |
| 6.2 Σημασία για Εφαρμογές και Νευροεπιστήμη   | 96         |
| 6.3 Περιορισμοί και Επιφυλάξεις   | 97         |
| 6.4 Προοπτικές Μελλοντικής Έρευνας  | 97         |
| 6.5 Επίλογος  | 98         |
| Γλωσσάρι Συντομογραφιών   | 100        |
| Glossary of Abbreviations (English)   | 102        |
| Βιβλιογραφία  | 104        |
| Παραθέματα  | 119        |

Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

## Πίνακας Σχημάτων

| Σγήμα 2.1 Η τοποθέτηση των ηλεκτροδίων ΗΕΓ με βάση το διεθνές σύστημα 10-20 25  |
|---|
| Σγήμα 2.2 Οι ζώνες συγνοτήτων σε EEG  |
| $\Sigma$ γήμα 2.3 Morlet wavelet  |
| Σγήμα 2.4 Χρονική και συγνοτική εξομάλυνση. (α) Για ένα γρονικό παράθυρο σταθερού   |
| μήκους, η γρονική και η συγνοτική εξομάλυνση παραμένουν σταθερές. (β) Για γρονικά   |
| παράθυρα που μειώνονται με τη συγνότητα, η γρονική εξομάλυνση μειώνεται και η   |
| συγγοτική εξομάλυνση αυξάνεται με τη συγγότητα. Πηνή Fieltrip   |
| $\Sigma_{\rm r}$ Σγήμα 2.5 Head model με σωστη τοποθέτηση ηλεκτοοδίων (κόκκινα σημεία) Η  |
| $\Delta_{\mu}$ μμα 2.5 πεαα πισασι μο σωστη τοιχουστηση προκεροσιων (κοικινα σημοια). Η<br>τμηματοποίηση και η δημιουργία του μοντέλου έγινε με γρήση της μεθόδου BEM 35  |
| $\Sigma$ γήμα 2.6 Παράδεινμα εντοπισμού πυγής με γρήση της μεθόδου sLORETA 38   |
| $\Sigma_{\chi}$ ήμα 2.0 Παρασσιγμα στοιλισμου πηγής με χρηση της μεσόσου shorter π  |
| $\Sigma_{\lambda}$ ήμα 2.7 Ποκοστρωματικό στάτος από perceptions και συναρτησή perception   |
| $\Sigma_{\lambda}$ ήμα 2.0 Στακρήτη συνεικόνηση του ποοτεινόμενου πλαισίου. Ο αστερίσκος (*) συμβολίζει   |
| $2\lambda_{1}$ m $\beta_{1}$ $2\lambda_{1}$ m $\beta_{1}$ $\beta_{2}$ $\beta_{1}$ $\beta_{2}$ $\beta_{1}$ $\beta_{2}$ $\beta_{1}$ $\beta_{2}$ $\beta_{1}$ $\beta_{2}$ $\beta_{1}$ $\beta_{2}$ $\beta_{2}$ $\beta_{2}$ $\beta_{1}$ $\beta_{2}$ |
| 52<br>Στήμα 3.2 Οι επιδόσεις του ποοτεινόμενου μοντέλου όσον αφορά: (α) απώλεια εκπαίδευσης   |
| $2\lambda$ (μμα 5.2 Οι επισθοείς του προτεινόμενου μοντελού σουν αφορά. (α) απωλεία εκλαισευοίζει<br>νια τις στονές- (β) απώλεια στικύρωσης νια τις στονές- (ν) ναρακτηριστική καιμτύλη   |
| $f(\alpha + \alpha) = f(\beta) = \alpha + \alpha$   |
| $\Sigma_{\alpha}$   |
| $2\chi(\mu u 3.5 (u) \pi v u \kappa u \zeta 00)\chi(00) = 100 \pi potervo \mu v v e \kappa u 0 - (p) \kappa u \mu v v m v m v m v m v m v m v m v m v m$  |
| 150 εποχες (μπχε) και 500 εποχες (μαιζεντα). Τα πιο οκούρα χρωματά αντιπροσωπεύουν μια $54$   |
| Στόμα 2 4 Ένα παράδουμα αμαπαράστασης του τιμόυ SHAP. Οι είποδοι υποδηλόμουμ τους   |
| $2\chi_{\rm II}$ μα 5.4 EVa παρασειγμα αναπαραστασης των τιμών SHAP. Οι είσουσι υποσηλώνουν τους  |
| πινακές ΤΓΚ, ένω η αντιοτοιχία του μοντέλου αντικατοπτριζεί τις επιορασείς των $(100  μοντέλου αντικατοπτριζεί τις επιορασείς των)$   |
| χρωματισμένων περιοχών σε κάθε συγκεκριμένη έξοσο κλάσης. Η επάνω σειρά αντιστοιχεί   |
| $S_{1}$ στην κλαση Faligue ενώ η κατώ σειρα στην κλαση Reside   |
| 2χημα 4.1 Ροη πληροφοριών και ρασικά στοιχεία των εμπροσθίων και αντιστροφών  |
| προρληματών   |
| 2χημα 4.2 Παρασειγματά του τροπού με τον οποίο ο αριθμος των ηλεκτροσίων και των  |
| οιπολων (σημεια πηγης) επηρεαζουν την εξοδο ΕSI. Σε καθε πινακά ο υπολογισμενος   |
| πινακας Leadfield (θερμικός χαρτης) εξαρταται από τα καναλια ηλεκτροδιών και τα σημεία  |
| πηγης   |
| $2\chi$ ημα 5.1 Η προβολη των τιμών SHAP σε εγκεφαλικό φλοιο  |
| 2χημα 5.2 Οι τιμες SHAP για καθε μια απο τις κλάσεις, (a) "αριστερό", (b) "δεξί", (c) "πόδια"   |
| και (d) "γλωσσα"  |
| Σχήμα 5.3 Απεικόνιση Gradient SHAP στον εγκέφαλο για τη νοερή κίνηση δεξιού χεριού,   |
| στον χρόνο 0,348 s  |

## Λίστα Πινάκων

| Πίνακας 1 Οι επιδόσεις του προτεινόμενου μοντέλου                               | 54      |
|---|---------|
| Πίνακας 2 Σύγκριση κόστους-αποτελεσματικότητας αλγορίθμων                       | 55      |
| Πίνακας 3 Σύγκριση μεθόδων ταξινόμησης ψυχικής κόπωσης                          | 58      |
| Πίνακας 4 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση αριστερού χεριού και η  | γνωστή  |
| συνάφειά τους   |         |
| Πίνακας 5 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση δεξιού χεριού και η     | γνωστή  |
| συνάφειά τους   |         |
| Πίνακας 6 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση ποδιών και η γνωστή σι  | ονάφειά |
| τους  |         |
| Πίνακας 7 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση γλώσσας και η γνωστή σι | ονάφειά |
| τους  | 88      |

Τεχνικές Βελτίωσης Επεξηγησιμότητας Συστημάτων Τεχνητής Νοημοσύνης σε Δεδομένα Ηλεκτροεγκεφαλογραφήματος

### Extended Summary: Techniques for Improving the Explainability of Artificial Intelligence Systems in Electroencephalography Data

This PhD thesis, titled "Techniques for Improving the Explainability of Artificial Intelligence Systems in Electroencephalography Data," addresses the critical challenge of enhancing the interpretability of advanced Artificial Intelligence (AI), particularly Deep Learning (DL), models applied to Electroencephalography (EEG) data. While DL models have demonstrated remarkable performance in classifying EEG signals for various applications, including Brain-Computer Interfaces (BCIs) and mental state detection, their inherent "black box" nature often hinders clinical adoption and scientific understanding. This research aims to bridge the gap between high-performance AI and the need for physiologically valid, anatomically informed explanations. The work is structured around three main axes: leveraging time-frequency analysis for explainable fatigue detection, critically reviewing EEG source localization techniques, and finally, integrating source localization with explainable AI for motor imagery BCI systems.

### Chapter 1: Introduction

Electroencephalography (EEG) is a fundamental non-invasive neurophysiological monitoring technique that records the brain's electrical activity with high temporal resolution. Its clinical value is undisputed in diagnosing and managing neurological conditions like epilepsy and sleep disorders, as well as monitoring brain function in critical states. Concurrently, EEG plays a central role in the development of Brain-Computer Interfaces (BCIs), which translate brain activity into control commands for external devices, offering new communication and interaction possibilities, especially for individuals with severe motor disabilities.

Despite the rich information in EEG signals, their analysis is challenging due to signal complexity, variability, and susceptibility to noise and artifacts. Artificial Intelligence (AI), particularly Deep Learning (DL) architectures, has shown significant success in automatic feature extraction and classification of EEG signals, achieving high performance in applications like mental fatigue detection or intention recognition in BCIs. However, the inherent complexity of these models often leads to a lack of transparency in their decision-making processes, rendering them "black boxes". This opacity is a major obstacle to their widespread adoption in clinical settings, where understanding the rationale behind a diagnosis or prediction is paramount for trust, regulatory compliance, and further scientific discovery.

The core problem this thesis addresses is the limited explainability of AI models when operating directly on sensor-level EEG data. This approach often fails to provide physiologically interpretable explanations due to the spatial blurring caused by volume conduction and the lack of direct mapping to specific brain regions. To tackle this, the research focuses on utilizing pre-processed EEG features that can imbue physiological meaning into the inputs of DL models. Two main feature extraction avenues are explored: time-frequency analysis, which reveals dynamic changes in brain rhythm power, and EEG source localization, which estimates the neural origins of brain activity, allowing its projection onto specific

anatomical brain regions. The working hypothesis is that integrating such physiologically informed features with modern Explainable Artificial Intelligence (XAI) techniques can significantly improve the interpretability of model predictions, making decision-making processes more transparent and understandable.

The primary objective of this thesis is to develop and evaluate a comprehensive framework that combines advanced EEG signal processing techniques, DL models, and XAI methods to enhance explainability in EEG classification for mental fatigue detection and BCI systems. Specifically, the research aims to apply and compare time-frequency analysis and source localization methods, including brain parcellation into anatomically defined regions, as pre-processing stages for DL models. Subsequently, XAI techniques are applied to attribute model predictions to specific input features, whether they are time-frequency components or activity in specific brain regions. This approach is expected to provide clearer and more neurophysiologically relevant explanations, bridging the gap between the computational power of DL models and the need for transparent, reliable results in clinical practice and research.

The thesis is structured with Chapter 2 presenting the theoretical background on EEG principles, signal processing methodologies such as time-frequency analysis and source localization, fundamentals of machine and deep learning with an emphasis on specific network types, and core concepts of XAI. Chapter 3 details the first research study, focusing on applying time-frequency analysis for mental fatigue detection and analyzing model explainability. Chapter 4 presents a published review of current EEG source localization techniques, their limitations, and future challenges. Chapter 5 describes the third research study, which implements a source localization approach combined with brain parcellation for a BCI system, investigating how this pre-processing affects the performance and explainability of DL models. Finally, Chapter 6 summarizes the main conclusions, discusses the research contributions, and proposes directions for future work.

#### Chapter 2: Prerequisites and Modern Techniques

This chapter lays the groundwork for the thesis by detailing the fundamental concepts and techniques employed in the analysis of EEG data and the application of AI.

EEG is a non-invasive method for recording brain electrical signals from the scalp, characterized by its high temporal resolution, which allows detection of rapid cortical dynamic changes in the order of milliseconds. This immediacy contrasts with methods like fMRI, which track slower blood flow changes. The relative affordability and portability of EEG equipment facilitate its broad application in various fields including neurology and BCIs. EEG systems utilize different types of electrodes, primarily wet electrodes, which use conductive gel for high signal quality, and dry electrodes, which offer convenience but potentially lower signal quality. This thesis primarily employed high-density wet EEG. Standardized electrode placement systems like the International 10-20 system are crucial for consistency, labeling electrodes based on cortical locations (e.g., F for Frontal, C for Central).

EEG signals are analyzed into distinct frequency bands, each associated with different neurophysiological states: Delta (0.5-3.5 Hz) with deep sleep; Theta (3.5-7 Hz) with drowsiness and working memory load; Alpha (7-15 Hz) with relaxed wakefulness; Beta (15-

30 Hz) with active thinking and motor planning; and Gamma (30-70 Hz) with higher-level cognitive processes and feature binding. The analysis of power and phase of these rhythms provides insights into brain function. However, EEG is susceptible to artifacts from environmental and physiological sources (EOG, ECG, EMG), necessitating careful artifact correction and preprocessing. This involves creating controlled recording environments, participant preparation, and computational techniques such as resampling, referencing, digital filtering (using FIR or IIR filters like Chebyshev), detrending, ICA for artifact separation, and baseline correction for event-related analyses.

Time-Frequency Analysis (TFA) is essential for understanding the dynamic interplay of time and frequency in EEG. Wavelet transforms, particularly Morlet wavelets, are effective for analyzing non-stationary EEG signals by providing a time-frequency representation with good resolution in both domains. This method allows for the extraction of features like Power Spectral Density (PSD) and phase information across different frequency bands.

EEG Source Localization (ESL) techniques aim to identify the brain regions generating scalprecorded EEG signals. This involves solving the "inverse problem," which is ill-posed due to the limited number of sensors compared to neural sources. Solving this requires a solution to the "forward problem"—calculating scalp potentials from known sources—which depends on an accurate head model representing tissue properties. Realistic head models are often created from MRI data using methods like the Boundary Element Method (BEM) or Finite Element Method (FEM). The Leadfield Matrix (LFM) then quantifies the contribution of each potential brain source to the measured EEG signals. Various inverse solution methods exist, including dipole fitting methods, distributed source models like Minimum Norm Estimate (MNE) and LORETA and its variants, and adaptive spatial filtering techniques like beamformers (e.g., LCMV). More recent approaches also incorporate machine learning and Bayesian methods.

Machine Learning (ML) and Deep Learning (DL) are extensively used for EEG analysis. Traditional ML involves manual feature engineering followed by classifiers like SVMs. DL, a subset of ML, uses multi-layered neural networks to automatically learn hierarchical data representations. Convolutional Neural Networks (CNNs) are particularly suited for EEG, automatically learning spatio-temporal features. Training techniques for DL models include backpropagation, regularization methods like dropout and batch normalization, and activation functions like softmax for output probabilities. Model performance is evaluated using metrics such as accuracy, sensitivity (recall), specificity, precision, F1-score, and AUROC.

Given the "black box" nature of many DL models, Explainable AI (XAI) tools are crucial for interpreting their decisions. Post-hoc XAI methods like LIME, gradient-based saliency maps, Integrated Gradients (IG), and SHAP (SHapley Additive exPlanations) are used. SHAP, based on game theory, assigns an importance value to each input feature for a specific prediction and offers a unified framework for interpretation. The GradientExplainer method within SHAP is particularly suited for DL models, efficiently estimating feature contributions.

#### Chapter 3: Explainability using Time-Frequency Analysis – Mental Fatigue Detection

This chapter presented a study focused on detecting mental fatigue from EEG signals, employing a computationally efficient shallow Convolutional Neural Network (CNN). The

approach combined time-frequency analysis for robust feature extraction with SHAP values for model explainability, aiming for high accuracy and real-time applicability with interpretable outcomes. Mental fatigue, often resulting from sleep deprivation, impairs cognitive functions, particularly working memory, which is critical in high-stakes environments such as healthcare. While traditional EEG analysis methods for fatigue detection exist, they often require manual intervention or lack generalization, and complex DL models can be computationally demanding and opaque.

The study involved 22 medical staff (20 after exclusions) who performed a visual n-back working memory task before ("Rested") and after ("Fatigued") on-call hospital shifts. EEG data was acquired using a 64-channel system, preprocessed (including downsampling, filtering, ICA for artifact correction, segmentation into 3-second epochs, and baseline correction), and then transformed using Morlet wavelets to generate Time-Frequency Representation (TFR) maps (63 channels×40 frequencies×13 time bins) for each trial. These TFR maps served as input to a shallow CNN architecture consisting of three 2D convolutional layers and one fully connected layer, with batch normalization and dropout for regularization. The model was trained using a subject-combined strategy to foster the learning of generalized fatigue features and achieved convergence in about 150 epochs on a CPU.

The proposed model demonstrated excellent performance, achieving 97% accuracy and a 97% F1-score in distinguishing between fatigued and rested states. The training duration was 44 minutes, with a prediction time of approximately 4 ms per instance, highlighting its efficiency and suitability for real-time applications. A comparative analysis showed that this approach was more computationally and time-efficient than several other contemporary AI models for fatigue detection when re-implemented under the same conditions.

To understand the model's decision-making process, SHAP values were calculated using the DeepLIFT method via the DeepExplainer tool. The analysis revealed that TFR power in the 5-15 Hz range (encompassing theta and alpha bands) contributed positively to the prediction of the "Fatigued" class. This finding is consistent with existing literature that associates alterations in theta (4-7 Hz) and alpha (8-12 Hz) activity with increased working memory load and mental fatigue. Conversely, activity in the 3-7 Hz band often indicated an inhibitory factor for the "Fatigued" class or a positive indicator for the "Rested" class, potentially reflecting memory load interactions with theta waves and task-related reactive control.

In discussion, the study underscored the efficacy of using a shallow CNN with engineered TFR features for high-accuracy, computationally inexpensive fatigue detection. The explainability analysis confirmed the model's reliance on physiologically relevant EEG patterns, enhancing its trustworthiness. The research acknowledged limitations such as the modest sample size for DL standards, although comparable to similar studies, and the operational definition of fatigue. Future work includes expanding datasets, incorporating diverse cognitive load measures, and further refining DL architectures and XAI techniques for even deeper insights into the neural underpinnings of mental fatigue.

#### Chapter 4: Modern EEG Source Localization Techniques: A Review

This chapter, derived from a published review by the author, provides a thorough examination of current methodologies in EEG Source Localization (ESL), also referred to as EEG Source Imaging (ESI). ESL techniques aim to identify the anatomical origins of electrical activity

recorded from the scalp, thereby offering insights into brain function by integrating temporal and spatial information from EEG signals. While EEG boasts superior temporal resolution over methods like fMRI, the process of localizing these sources is inherently an "ill-posed problem" because the number of recording electrodes is vastly outnumbered by the potential neural generators.

The core of ESL revolves around solving two interconnected problems: the forward problem and the inverse problem. The forward problem involves calculating the electric potentials that would appear on the scalp given a known distribution of electrical sources within the brain and a conductive head model. The accuracy of this calculation, often based on Poisson's equation, is paramount and relies heavily on a precise model of the head's geometry and the electrical conductivity of its various tissues (scalp, skull, cerebrospinal fluid (CSF), and brain tissue). While individual MRI scans provide the best anatomical data for these models, template MRIs are often used when subject-specific scans are unavailable. Head models can range from simpler spherical models to more realistic ones derived using the Boundary Element Method (BEM) or the Finite Element Method (FEM), with FEM allowing for more complex tissue properties like anisotropy but at a higher computational cost. The relationship between sources and scalp potentials is encapsulated in the Leadfield Matrix (LFM).

The inverse problem, conversely, is the process of estimating the locations and strengths of the neural sources from the recorded scalp EEG data. Due to its underdetermined nature, the inverse problem does not have a unique solution without the imposition of a priori constraints, which can be mathematical, anatomical, or neurophysiological, to limit the space of possible solutions. Various methods have been developed to tackle the inverse problem. These broadly fall into categories such as dipole fitting methods, which model brain activity as arising from a few discrete current dipoles (ECDs), and distributed source models, which assume that activity is spread across a grid or volume within the brain. Popular distributed source methods include the Minimum Norm Estimate (MNE) and its variations (e.g., dw-MNE), as well as LORETA (Low-Resolution Electromagnetic Tomography) and its refined versions like sLORETA and eLORETA, which typically enforce some form of spatial smoothness on the estimated source distribution. Another important class of methods is beamformers, such as the Linearly Constrained Minimum Variance (LCMV) beamformer, which use adaptive spatial filters to estimate activity from specific brain locations while suppressing interference from other sources. More recently, machine learning techniques, including deep neural networks (e.g., ConvDip, DST-DAE) and sophisticated Bayesian approaches incorporating statistical priors, have shown promise in improving the accuracy and efficiency of ESL.

ESL has found diverse applications in both clinical settings and cognitive research. Clinically, it is instrumental in the pre-surgical evaluation of epilepsy patients for localizing epileptogenic zones, aiding in the diagnosis and monitoring of neurodegenerative diseases like Alzheimer's disease through source-space connectivity analysis, studying sleep disorders by identifying regions involved in different sleep stages or parasomnias, and supporting neurorehabilitation efforts by decoding movement intentions in stroke patients. In cognitive neuroscience, ESI helps in localizing ERP components associated with specific cognitive functions (e.g., P300 for attention), understanding the neural basis of face and emotion processing, and investigating brain connectivity during tasks involving language or working memory.

Despite significant advancements, ESL faces ongoing challenges. These include its sensitivity to noise and artifacts which can lead to localization errors or "ghost" sources, the fundamental ill-posed nature of the inverse problem affected by electrode density versus the number of potential dipoles (Figure 4.2 illustrates this trade-off), inaccuracies in electrode corregistration with the head model, and the critical dependence on accurate tissue conductivity values and head model realism. Newer machine learning methods also face hurdles such as the lack of large, accurately labeled real-world ESI datasets for training, while complex Bayesian methods can be computationally prohibitive for real-time applications. Future directions in ESL point towards combining data-driven and model-based approaches, further improving computational efficiency and localization accuracy, integrating information from multiple imaging modalities (e.g., EEG-fMRI fusion), and enhancing the robustness of these techniques for broader clinical and research applicability.

# Chapter 5: Explainability using Source Localization Technique in a Brain-Computer Interface (BCI)

This pivotal chapter details the development and evaluation of a framework aimed at enhancing the interpretability of neural networks used in motor imagery (MI) based Brain-Computer Interfaces (BCIs). The core idea is to transcend the limitations of sensor-level EEG analysis by transforming EEG data into source space and then parcellating this activity into anatomically defined brain regions before feeding it to a classifier, thereby making the model's explanations more aligned with neuroanatomical knowledge. Traditional BCI approaches relying on scalp EEG often struggle with low spatial resolution due to volume conduction effects, making it challenging to accurately identify the cortical areas driving the BCI and to derive meaningful interpretations from explainability maps.

The proposed methodology integrates several advanced techniques. First, EEG source localization using an LCMV (Linearly Constrained Minimum Variance) beamformer is employed to project the scalp-recorded EEG signals into the brain volume, thus estimating the neural origins of the observed activity. This source-reconstructed activity is then combined with an anatomical parcellation scheme, specifically the Desikan-Killiany atlas, which divides the cortex into 68 distinct Regions of Interest (ROIs). By averaging the source activity within each ROI, a set of meaningful regional time series is generated for each EEG trial. These parcellated source-space EEG features subsequently serve as input to a customdesigned 1D Convolutional Neural Network (CNN1D\_Custom) for the classification of four different motor imagery tasks (left hand, right hand, feet, and tongue). Finally, various Explainable AI (XAI) techniques, including Saliency Maps, Integrated Gradients (IG), and notably SHAP (SHapley Additive exPlanations), are applied to interpret the decisions of the trained CNN in terms of the contributions of these anatomically defined brain regions. The central hypothesis is that this pipeline will yield explanations that are not only more accurate in terms of neuroanatomical validity but also more readily interpretable, thereby bridging the gap between computational models and clinical or neuroscientific understanding. This approach is presented as a novel contribution, as previous work on BCI explainability has predominantly focused on sensor-space interpretations.

The methodology was evaluated using the BCI Competition IV dataset 2a, a standard benchmark for MI BCIs, comprising EEG data from 9 subjects performing the four MI tasks. Preprocessing steps included ICA for artifact removal, band-pass filtering (8-30 Hz) to capture relevant mu and beta rhythms, epoching into 2-second windows aligned with MI cues,

and baseline correction. For source localization, an MNI 152 template brain and a 3-shell BEM head model were utilized to compute the leadfield matrix, from which LCMV spatial filters were derived and applied to generate "virtual sensor" time series at numerous cortical locations. These source activities were then averaged within each of the 68 Desikan-Killiany ROIs, resulting in a 68 regions×500 time points input matrix for the CNN1D\_Custom model for each trial. The CNN architecture itself consisted of two 1D convolutional layers followed by max-pooling, then flattening and passing through two dense layers before a final softmax output layer (detailed in Code 1 of the thesis appendix ). The model was trained subject-specifically using categorical cross-entropy loss and the Adam optimizer, achieving a competitive classification accuracy of 76%.

The core results stemmed from the XAI analysis, particularly using SHAP values, to determine which brain regions' activities were most influential for the CNN's classification of each MI task. For left-hand MI, the Right Precentral Gyrus (R M1) and Right Postcentral Gyrus (R S1) exhibited the largest positive SHAP values, indicating their activity was proclassification, consistent with contralateral motor control. Conversely, the ipsilateral Left Precentral and Postcentral Gyri showed strong negative SHAP values (anti-classification) (Table 4). A mirrored pattern was observed for right-hand MI, with the Left Precentral Gyrus (L M1) and Left Postcentral Gyrus (L S1) showing maximal positive SHAP values, and the ipsilateral right-hemisphere motor areas showing negative contributions (Table 5, Figure 5.3 ). For foot MI (both feet), the Paracentral Lobules in both hemispheres (representing foot/leg areas) and the medial Superior Frontal Gyrus (SMA) displayed the highest positive SHAP values, while lateral motor areas associated with hand movements contributed negatively (Table 6 ). Finally, for tongue MI, bilateral activation of the lateral Precentral Gyri (tongue/face M1 areas), along with the Left Inferior Frontal Gyrus (Broca's area) and Left Postcentral Gyrus (face S1 area), were identified as key positive contributors, reflecting the more bilateral control of tongue movements and potential engagement of speech-related networks (Table 7). Figure 5.1 and Figure 5.2 provide visualizations of these SHAP value distributions.

The discussion of these results highlighted the strong congruence between the SHAPidentified important regions and the well-established neuroscientific understanding of motor system organization, effectively reproducing a functional "homunculus". This confirmed that the CNN learned to rely on physiologically genuine neuroanatomical patterns rather than spurious features. The approach successfully captured laterality effects for hand movements and offered enhanced interpretability compared to sensor-level analyses by directly linking model decisions to specific anatomical ROIs, resolving ambiguities inherent in scalp EEG. The ability to identify contributions from areas like Broca's region in tongue MI, which might be obscured in sensor-space, further underscored the value of this source-based explainability.

Despite these successes, limitations were acknowledged, including the inherent inaccuracies in source localization with template models, the focus on sustained MI activity potentially missing finer temporal dynamics, the assumptions of XAI methods like SHAP regarding feature independence in the context of correlated brain signals, and the restriction of the analysis to cortical regions due to EEG's poor sensitivity to deep brain structures like the cerebellum or basal ganglia. Future research directions include enhancing spatial/temporal resolution, integrating neurobiological priors into models, exploring feature interactions, developing anatomically informed neurofeedback strategies, and extending the framework to clinical populations and other BCI paradigms.

#### Chapter 6: Discussion and Epilogue

This concluding chapter synthesizes the findings and contributions of the dissertation, reflects on their significance for AI applications in EEG and neuroscience, acknowledges limitations, and outlines promising avenues for future research. The overarching aim of the thesis was to improve the transparency and interpretability of AI models applied to EEG data, with a particular focus on BCI systems. This was achieved by systematically investigating the roles of time-frequency analysis, EEG source localization, and explainable deep learning techniques.

The research first demonstrated, in Chapter 3, that employing time-frequency analysis as a feature engineering step allows even shallow CNNs to achieve high accuracy (97%) in complex tasks like mental fatigue detection, with the added benefits of computational efficiency and interpretable results via SHAP analysis, which highlighted physiologically relevant theta/alpha band activity. This established the value of carefully chosen input representations for building efficient and explainable AI systems suitable for real-time applications, including EDGE-AI scenarios.

Chapter 4 provided a critical review of EEG source localization methods, delineating their theoretical foundations, algorithmic variations (such as beamformers and minimum norm solutions), diverse applications, and inherent challenges (e.g., solving the forward and inverse problems, head model accuracy, sensitivity to deep sources). This review underscored the necessity of transitioning from sensor-level to source-level analysis to achieve anatomically grounded interpretations, thereby justifying the use of beamformer-based source localization in the subsequent empirical work.

The central contribution, detailed in Chapter 5, was the development and validation of an integrated framework that combines source-localized and anatomically parcellated EEG data with a CNN classifier and XAI methods for a motor imagery BCI. This approach yielded competitive classification accuracy while, more importantly, producing explanations (primarily through SHAP values) that mapped directly onto specific cortical regions of interest. The XAI results robustly demonstrated that the model's decisions were based on activation patterns consistent with the known somatotopic organization of the motor cortex— a functional "homunculus"—and accurately captured laterality effects associated with hand movements. This direct linkage of AI model behavior to neuroanatomical structures significantly demystifies the "black box," offering a level of transparency rarely achieved with sensor-space EEG analyses. The ability to pinpoint contributions from specific regions, such as Broca's area during tongue imagery, further illustrates the power of this source-based explainability to reveal subtle neural correlates.

The implications of these findings are substantial for both clinical applications and fundamental neuroscience. For clinical BCIs, particularly in neurorehabilitation, the ability to verify that the AI model relies on expected, physiologically relevant brain activity (e.g., in the motor cortex) enhances trust and facilitates the design of targeted, anatomically informed neurofeedback protocols. It also provides a means to monitor neuroplastic changes during recovery. For neuroscience, this framework offers a powerful tool to validate existing

knowledge and to generate new hypotheses about brain function by revealing data-driven associations between brain activity in specific regions and cognitive tasks, such as the observed engagement of Broca's area during tongue MI. This fosters a closer dialogue between AI development and neuroscientific inquiry.

However, several limitations are acknowledged. The precision of EEG source localization, while improved, remains finite, especially when using template head models and atlases rather than individual anatomical data. The datasets used, particularly for the BCI experiments, were of moderate size for deep learning standards, and generalization to larger, more diverse populations needs further investigation. The inherent insensitivity of scalp EEG to deep brain structures means the current analysis was largely confined to cortical activity. Furthermore, the temporal dynamics of brain activity were somewhat simplified in the explainability analyses, and the XAI methods themselves have underlying assumptions that might be tested by the complex, correlated nature of brain signals.

These limitations naturally lead to several promising avenues for future research. These include efforts to further improve the spatial and temporal resolution of the EEG analysis pipeline through higher-density EEG systems, individualized head models, and more refined anatomical or functional parcellation schemes. Developing XAI methods that can capture more complex temporal dynamics and inter-regional interactions is another key direction. The integration of multimodal neuroimaging data (e.g., concurrent EEG-fMRI) could provide a more comprehensive understanding of the neural networks involved, including subcortical contributions. Incorporating neuroscientific priors directly into the architecture or training of AI models could also enhance their performance, robustness, and inherent interpretability. Crucially, applying and validating this framework in diverse clinical populations (e.g., stroke patients) will be essential for translating these research advancements into tangible benefits for patients by allowing for the personalization of BCI and neurofeedback therapies. Finally, the versatility of the proposed framework suggests its potential extension to a wide array of other EEG-based classification tasks, from assessing cognitive load and emotional states to detecting sleep stages or pathological brain activity like epileptic seizures, thereby broadly contributing to the quest for more understandable and trustworthy AI in brain research and clinical neurology.

In summary, this dissertation has presented a cohesive and innovative framework that successfully integrates advanced signal processing, deep learning, and state-of-the-art explainability techniques to create high-performing EEG analysis models that are also anatomically interpretable. By demonstrating, through a motor imagery BCI application, that the often-competing goals of accuracy and explainability can be reconciled, this work marks a significant step forward. The capacity to link AI model decisions to specific brain regions in a manner consistent with established neuroscientific knowledge is paramount for fostering clinical adoption and advancing our fundamental understanding of the brain. The proposed methodology helps lay the groundwork for a new generation of brain-computer interfaces and neuroscientific tools where a deep understanding of the underlying neurophysiology is integral to both design and application, fostering a transparent and synergistic collaboration between humans and artificial intelligence. This vision ultimately aims to accelerate the translation of EEG-based technologies from research laboratories into routine clinical practice and everyday life, for the benefit of both patients and users.

### 1. Κεφάλαιο 1 - Εισαγωγή

Η παρούσα διδακτορική διατριβή διερευνά μεθόδους για την ενίσχυση της επεξηγησιμότητας των μοντέλων βαθιάς μάθησης που εφαρμόζονται σε προβλήματα ταξινόμησης ηλεκτροεγκεφαλογραφικών (ΗΕΓ) σημάτων. Το ΗΕΓ αποτελεί μια θεμελιώδη μη επεμβατική τεχνική νευροφυσιολογικής παρακολούθησης, καταγράφοντας την ηλεκτρική δραστηριότητα του εγκεφάλου με υψηλή χρονική ανάλυση [1]. Η κλινική του αξία είναι αδιαμφισβήτητη στη διάγνωση και διαχείριση νευρολογικών παθήσεων, όπως η επιληψία και οι διαταραχές ύπνου, καθώς και στην παρακολούθηση της εγκεφαλικής λειτουργίας σε κρίσιμες καταστάσεις. Παράλληλα, το ΗΕΓ διαδραματίζει κεντρικό ρόλο στην ανάπτυξη Συστημάτων Διεπαφής Εγκεφάλου-Υπολογιστή (Brain-Computer Interfaces - BCIs), τα οποία μεταφράζουν την εγκεφαλική δραστηριότητα σε εντολές ελέγχου εξωτερικών συσκευών, προσφέροντας νέες δυνατότητες επικοινωνίας και αλληλεπίδρασης, ιδιαίτερα σε άτομα με σοβαρές κινητικές αναπηρίες [2].

Η πολυπλοκότητα και η μεταβλητότητα των ΗΕΓ σημάτων, σε συνδυασμό με την ευαισθησία τους σε θόρυβο και τεχνουργήματα, καθιστούν την ανάλυσή τους μια σημαντική πρόκληση. Η Τεχνητή Νοημοσύνη (TN), και ειδικότερα οι αρχιτεκτονικές βαθιάς μάθησης, έχουν επιδείξει αξιοσημείωτη επιτυχία στην αυτόματη εξαγωγή χαρακτηριστικών και στην ταξινόμηση των ΗΕΓ σημάτων, επιτυγχάνοντας υψηλές επιδόσεις σε εφαρμογές όπως η ανίχνευση νοητικής κόπωσης ή η αναγνώριση προθέσεων σε συστήματα BCI [3]. Ωστόσο, η εγγενής πολυπλοκότητα αυτών των μοντέλων συχνά οδηγεί σε έλλειψη διαφάνειας ως προς τον τρόπο λήψης των αποφάσεών τους, καθιστώντας τα «μαύρα κουτιά» [4]. Αυτή η αδιαφάνεια αποτελεί σημαντικό εμπόδιο για την ευρεία υιοθέτησή τους σε κλινικά περιβάλλοντα, όπου η κατανόηση της αιτιολόγησης πίσω από μια διάγνωση ή πρόβλεψη είναι υψίστης σημασίας για την εμπιστοσύνη, την κανονιστική συμμόρφωση και την περαιτέρω επιστημονική ανακάλυψη [5].

Το κεντρικό θέμα που η παρούσα διατριβή προσπαθεί να αντιμετωπίσει είναι η περιορισμένη επεξηγησιμότητα των μοντέλων ΤΝ όταν αυτά λειτουργούν απευθείας στα δεδομένα ΗΕΓ που καταγράφονται από τους αισθητήρες στο τριχωτό της κεφαλής. Αυτή η προσέγγιση συχνά αποτυγχάνει να παρέχει φυσιολογικά ερμηνεύσιμες εξηγήσεις λόγω της χωρικής ασάφειας που προκαλείται από το φαινόμενο της αγωγιμότητας όγκου και της έλλειψης άμεσης αντιστοίχισης με συγκεκριμένες εγκεφαλικές περιοχές. Για την αντιμετώπιση αυτού του προβλήματος, η έρευνα επικεντρώνεται στην αξιοποίηση προεπεξεργασμένων χαρακτηριστικών από τα σήματα ΗΕΓ, τα οποία μπορούν να προσδώσουν φυσιολογική σημασία στις εισόδους των μοντέλων βαθιάς μάθησης. Συγκεκριμένα, διερευνώνται δύο κύριες κατευθύνσεις εξαγωγής γαρακτηριστικών: η ανάλυση γρόνου-συγνότητας, η οποία αποκαλύπτει τις δυναμικές αλλαγές στην ισγύ των εγκεφαλικών ρυθμών, και ο εντοπισμός πηγής, ο οποίος εκτιμά τις νευρωνικές πηγές της εγκεφαλικής δραστηριότητας, επιτρέποντας την προβολή της δραστηριότητας σε συγκεκριμένες ανατομικές περιοχές του εγκεφάλου. Η υπόθεση εργασίας είναι ότι η ενσωμάτωση τέτοιων φυσιολογικά πληροφορημένων χαρακτηριστικών, σε συνδυασμό με σύγχρονες τεχνικές Επεξηγήσιμης Τεχνητής Noημοσύνης (Explainable Artificial Intelligence - XAI), μπορεί να βελτιώσει σημαντικά την ερμηνευσιμότητα των προβλέψεων των μοντέλων, καθιστώντας τις διαδικασίες λήψης αποφάσεων πιο διαφανείς και κατανοητές.

Πρωταρχικός στόχος της παρούσας διατριβής είναι η ανάπτυξη και αξιολόγηση ενός ολοκληρωμένου πλαισίου που συνδυάζει προηγμένες τεχνικές επεξεργασίας ΗΕΓ σημάτων, μοντέλα βαθιάς μάθησης και μεθόδους ΧΑΙ, με σκοπό την ενίσχυση της επεξηγησιμότητας

στην ταξινόμηση ΗΕΓ για την ανίχνευση νοητικής κόπωσης και για συστήματα BCI. Ειδικότερα, η έρευνα στοχεύει στην εφαρμογή και σύγκριση μεθόδων χρόνου-συχνότητας ανάλυσης και εντοπισμού πηγής, συμπεριλαμβανομένης της κατάτμησης (parcellation) του εγκεφάλου σε ανατομικά καθορισμένες περιοχές, ως στάδια προεπεξεργασίας για την τροφοδότηση μοντέλων βαθιάς μάθησης. Ακολούθως, επιδιώκεται η εφαρμογή τεχνικών XAI για την απόδοση των προβλέψεων των μοντέλων σε συγκεκριμένα χαρακτηριστικά εισόδου, είτε αυτά αφορούν συνιστώσες χρόνου-συχνότητας είτε δραστηριότητα συγκεκριμένων εγκεφαλικών περιοχών. Η ελπίδα είναι ότι αυτή η προσέγγιση θα παρέχει σαφέστερες και νευροφυσιολογικά πιο συναφείς εξηγήσεις, γεφυρώνοντας το χάσμα μεταξύ της υπολογιστικής ισχύος των μοντέλων βαθιάς μάθησης και της ανάγκης για διαφανή και αξιόπιστα αποτελέσματα στην κλινική πράξη και την έρευνα.

Η δομή της διατριβής έχει ως εξής: Το Κεφάλαιο 1, η παρούσα εισαγωγή, περιγράφει το ιστορικό, τα κίνητρα, το πρόβλημα που εξετάζεται, τους στόχους και τη συνολική δομή της έρευνας. Το Κεφάλαιο 2 παρουσιάζει το θεωρητικό υπόβαθρο, καλύπτοντας τις βασικές αρχές του ΗΕΓ, τις μεθοδολογίες επεξεργασίας σήματος, την ανάλυση χρόνου συχνότητας (time-frequency analysis), τις θεμελιώδεις μεθόδους εντοπισμού πηγής και την κατάτμηση του εγκεφάλου. Επιπλέον, αναλύονται οι βασικές αρχές της μηχανικής και βαθιάς μάθησης, με έμφαση στα τροφοδοτικά και συνελικτικά νευρωνικά δίκτυα, καθώς και οι θεμελιώδεις έννοιες της επεξηγησιμότητας στην ΤΝ. Το Κεφάλαιο 3 περιγράφει την πρώτη ερευνητική εργασία της διατριβής, η οποία εστιάζει στην εφαρμογή time-frequency ανάλυσης για την ανίχνευση νοητικής κόπωσης και την ανάλυση της επεξηγησιμότητας των χρησιμοποιούμενων μοντέλων. Το Κεφάλαιο 4 παρουσιάζει μια δημοσιευμένη επισκόπηση σχετικά με τις τρέχουσες τεχνικές εντοπισμού πηγής ΗΕΓ, τους περιορισμούς τους και τις μελλοντικές προκλήσεις. Το Κεφάλαιο 5 αναφέρεται στην τρίτη ερευνητική εργασία, η οποία υλοποιεί μια προσέγγιση εντοπισμού πηγής σε συνδυασμό με κατάτμηση του εγκεφάλου για ένα σύστημα BCI, διερευνώντας πώς αυτή η προεπεξεργασία επηρεάζει την απόδοση και την επεξηγησιμότητα των μοντέλων βαθιάς μάθησης. Τέλος, το Κεφάλαιο 6 συνοψίζει τα κύρια συμπεράσματα της διατριβής, συζητά τη συνεισφορά της έρευνας και προτείνει κατευθύνσεις για μελλοντική εργασία.

### 2. Κεφάλαιο 2 – Προαπαιτούμενα και σύγχρονες τεχνικές

### 2.1 Ηλεκτροεγκεφαλογράφημα

Η ηλεκτροεγκεφαλογραφία (ΗΕΓ) είναι μία από τις πιο άμεσες, μη επεμβατικές προσεγγίσεις για την καταγραφή αυτών των ηλεκτρικών σημάτων σε επίπεδο τριχωτού της κεφαλής. Το κύριο πλεονέκτημά της είναι η υψηλή χρονική ανάλυση: Το ΗΕΓ μπορεί να ανιχνεύσει ταχείες μεταβολές στη δυναμική του φλοιού της τάξης των χιλιοστών του δευτερολέπτου, επιτρέποντας στους ερευνητές και τους κλινικούς γιατρούς να εξετάσουν πώς εξελίσσεται η εγκεφαλική δραστηριότητα σε πραγματικό χρόνο. Σε σύγκριση με άλλες μεθόδους νευροαπεικόνισης, όπως η λειτουργική απεικόνιση μαγνητικού συντονισμού (fMRI), η οποία μετρά αλλαγές στη ροή του αίματος σε δευτερόλεπτα, το ΗΕΓ προσφέρει ένα πιο άμεσο στιγμιότυπο της νευρωνικής δραστηριότητας. Επιπλέον, ο εξοπλισμός ΗΕΓ είναι σχετικά προσιτός και συμπαγής, διευκολύνοντας ευρείες κλινικές και ερευνητικές εφαρμογές, μεταξύ άλλων στη νευρολογία, την ψυχιατρική, την ψυχολογία και τα αναδυόμενα συστήματα διεπαφής εγκεφάλου-υπολογιστή (BCI).

### 2.1.1 Τύποι ηλεκτροδίων ΗΕΓ

Τα συστήματα ΗΕΓ βασίζονται σε ηλεκτρόδια που μπορούν να ταξινομηθούν σε δύο κύριες κατηγορίες:

- Υγρά ηλεκτρόδια. Αυτά τα ηλεκτρόδια αποτελούνται συνήθως από μικρούς δίσκους κατασκευασμένους από μέταλλα όπως ανοξείδωτο χάλυβα, κασσίτερο, χρυσό ή χλωριούχο άργυρο-αργύρου (Ag/AgCl). Για να βελτιωθεί η αγωγιμότητα μεταξύ του τριχωτού της κεφαλής και του ηλεκτροδίου, χρησιμοποιείται αγώγιμη γέλη ή πάστα. Παρά την καθιερωμένη αξιοπιστία και την υψηλή ποιότητα σήματος, τα υγρά ηλεκτρόδια απαιτούν προετοιμασία του δέρματος και εφαρμογή γέλης, η οποία μπορεί να είναι χρονοβόρα και λιγότερο άνετη για τους συμμετέχοντες [3].
- Ξηρά ηλεκτρόδια. Αντίθετα, τα ξηρά ηλεκτρόδια πραγματοποιούν άμεση μηχανική επαφή με το δέρμα, εξαλείφοντας την ανάγκη για αγώγιμα τζελ. Το μικρότερο μέγεθός τους και η πιο βολική εγκατάστασή τους έχουν δυνατότητες για πραγματικές και φορητές εφαρμογές [4]. Ωστόσο, τα τρέχοντα συστήματα ξηρών ηλεκτροδίων είναι συχνά περιορισμένα ως προς τον αριθμό των καναλιών (συνήθως έως 30) και ενδέχεται να παρουσιάζουν αυξημένη αντίσταση, οδηγώντας σε χαμηλότερη ποιότητα σήματος ή μεγαλύτερη ευαισθησία σε αντικείμενα κίνησης.

Για τους σκοπούς της παρούσας διατριβής, χρησιμοποιείται ΗΕΓ υγρών ηλεκτροδίων υψηλής πυκνότητας (64 κανάλια) για να εξασφαλιστεί εκτεταμένη κάλυψη των περιοχών του φλοιού και ισχυρή ποιότητα δεδομένων στη περίπτωση την εφαρμογής ανίχνευσης κόπωσης. Παρ' όλα αυτά, το ξηρό ΗΕΓ παραμένει μια πολύτιμη επιλογή σε σενάρια όπου η φορητότητα και η ταχύτητα εγκατάστασης είναι υψίστης σημασίας (π.χ. φορητό ΗΕΓ για παρακολούθηση στο σπίτι). Τα υποκείμενα εγκεφαλικά σήματα παραμένουν τα ίδια- είναι κυρίως η ποιότητα των δεδομένων, ο αριθμός των καναλιών και οι πρακτικές εκτιμήσεις που διαφέρουν μεταξύ των τύπων ηλεκτροδίων.

### 2.1.2 Τοποθέτηση ηλεκτροδίων ΗΕΓ και μοντάζ

Ένας κρίσιμος παράγοντας στην καταγραφή του ΗΕΓ είναι η τυποποιημένη τοποθέτηση των ηλεκτροδίων, εξασφαλίζοντας τη συνοχή μεταξύ των συμμετεχόντων, των συνεδριών και των εργαστηρίων. Δύο ευρέως αποδεκτές συμβάσεις είναι το **σύστημα 10-20** και το **σύστημα 10-10**. Αυτές οι ονοματολογίες αντικατοπτρίζουν τις ανάλογες αποστάσεις (10% ή 20%) μεταξύ των θέσεων των ηλεκτροδίων σε σχέση με το συνολικό μέγεθος του κεφαλιού.

### Το σύστημα 10-20

Στο σύστημα 10-20, τα ηλεκτρόδια επισημαίνονται για να δηλώσουν την κατά προσέγγιση θέση τους σε συγκεκριμένες περιοχές του φλοιού [5]. Κάθε ετικέτα υποδεικνύει τον αντίστοιχο λοβό του εγκεφάλου ή μια θέση στη μέση γραμμή:

- Προμετωπιαία (Fp) για ηλεκτρόδια κοντά στο μέτωπο
- Μετωπιαίο (F) για ηλεκτρόδια πάνω από τον μετωπιαίο λοβό
- Temporal (T) για ηλεκτρόδια πάνω από τον κροταφικό λοβό
- Κεντρικός (C) για ηλεκτρόδια κοντά στη μεσαία περιοχή του φλοιού (αν και ανατομικά δεν υπάρχει κεντρικός λοβός, ο συμβολισμός αυτός βοηθά στην κάλυψη μεταξύ μετωπιαίων και βρεγματικών περιοχών).
- Parietal (P) για ηλεκτρόδια πάνω από τον βρεγματικό λοβό
- Ινιακός (Ο) για ηλεκτρόδια πάνω από τον ινιακό λοβό
- To «z » υποδηλώνει τη μέση γραμμή (π.χ. Fz, Cz, Pz, Oz).
- Οι ζυγοί αριθμοί (π.χ. Fp2, F8, T6) αναφέρονται σε ηλεκτρόδια στη δεξιά πλευρά του τριχωτού της κεφαλής.
- Οι περιττοί αριθμοί (π.χ. Fp1, F7, T5) αναφέρονται σε ηλεκτρόδια στην αριστερή πλευρά του τριχωτού της κεφαλής.

Πρόσθετες ονοματολογίες όπως AF, FC, FT, CP, TP και PO επισημαίνουν ενδιάμεσες θέσεις μεταξύ των κύριων λοβών (π.χ., το AF είναι μεταξύ Fp και F, το CP είναι μεταξύ C και P). Τα μαστοειδή ηλεκτρόδια (M1, M2) τοποθετούνται ακριβώς πίσω από το αυτί και συχνά χρησιμεύουν ως αναφορές ή ως πρόσθετα κανάλια. Επιπλέον, η θέση inion (Iz) είναι σημαντική για την ευθυγράμμιση ορισμένων μοντάζ και βρίσκεται κοντά στην ινιακή προεξοχή [6].

Στην παρούσα εργασία (συγκεκριμένα, στα Κεφάλαια 3 και 5), τα δεδομένα ΗΕΓ έχουν καταγραφεί και ερμηνευτεί με τη χρήση του συστήματος 10-20, διασφαλίζοντας την ευθυγράμμιση των αποτελεσμάτων με την καθιερωμένη βιβλιογραφία και διευκολύνοντας την αναπαραγωγή τους από άλλους ερευνητές. Τέτοιες τυποποιημένες τοποθετήσεις ηλεκτροδίων επιτρέπουν μια πιο γενικευμένη ερμηνεία των δεδομένων, επιτρέποντας την κοινή ονοματολογία (π.χ. Fz, Cz, Pz) που αντιστοιχεί σε γνωστές φλοιώδεις περιοχές ή λειτουργικές περιοχές.

### Παραλλαγές και επεκτάσεις

Το σύστημα 10-10 και οι περαιτέρω επεκτάσεις (π.χ. σύστημα 10-5) παρέχουν μια πυκνότερη διάταξη τοποθέτησης ηλεκτροδίων για λεπτομερέστερη χωρική ανάλυση. Ωστόσο, για πολλά κλινικά και ερευνητικά σενάρια -συμπεριλαμβανομένων των περισσότερων μελετών

δυναμικού που σχετίζεται με γεγονότα (ERP) και BCI- το σύστημα 10-20 προσφέρει συνήθως μια επαρκή ισορροπία μεταξύ κάλυψης και πρακτικότητας, ειδικά όταν η εστίαση είναι σε ευρείες περιοχές του φλοιού και όχι σε εξαιρετικά εστιασμένες πηγές δραστηριότητας.



Σχήμα 2.1 Η τοποθέτηση των ηλεκτροδίων ΗΕΓ με βάση το διεθνές σύστημα 10-20

### 2.1.3 Ρυθμοί και ταλαντώσεις ΗΕΓ

Τα σήματα ΗΕΓ μπορούν να αναλυθούν σε ταλαντωτικές συνιστώσες, καθεμία από τις οποίες αντιστοιχεί σε ορισμένες περιοχές συχνοτήτων που συνδέονται με διακριτές νευροφυσιολογικές ή γνωστικές διεργασίες. Αυτοί οι **ρυθμοί** αντικατοπτρίζουν συχνά αλληλεπιδράσεις φλοιώδους δικτύου που διέπουν διάφορες καταστάσεις συνείδησης, εμπλοκής της προσοχής και γνωστικές εργασίες [7].

### 1. Δέλτα (δ: 0,5-3,5 Hz).

Τα κύματα δέλτα είναι οι συνιστώσες με τη χαμηλότερη συχνότητα και συνήθως συνδέονται με τον βαθύ ύπνο (ιδίως με τον ύπνο non-REM). Μπορούν επίσης να εμφανιστούν σε ορισμένες παθολογικές καταστάσεις ή σε εργασίες που απαιτούν σημαντικούς πόρους προσοχής σε ορισμένα άτομα.

#### 2. Θήτα (θ: 3,5-7 Hz).

Τα κύματα Θήτα εμφανίζονται συχνά κατά τη διάρκεια της υπνηλίας ή του ελαφρού ύπνου. Στους ενήλικες, η δραστηριότητα θήτα στο μετωπιαίο-μεσολόβιο έχει συνδεθεί με το φορτίο της εργαζόμενης μνήμης και τον γνωστικό έλεγχο, υποδηλώνοντας εμπλοκή σε σύνθετα νοητικά καθήκοντα και καταστάσεις αυξημένης εσωτερικής επεξεργασίας.

#### 3. Άλφα (α: 7-15 Hz).

Η δραστηριότητα άλφα συνδέεται κλασικά με μια χαλαρή αλλά και σε εγρήγορση κατάσταση, η οποία παρατηρείται πιο έντονα στις ινιακές περιοχές όταν τα μάτια είναι

κλειστά. Συχνά θεωρείται ένας ρυθμός «αδράνειας». Οι μεταβολές στην ισχύ άλφα μπορούν να αναδείξουν τις αλλαγές προσοχής, τη νοητική κόπωση και τις ατομικές διαφορές στη γνωστική απόδοση.

### 4. Βήτα (β: 15-30 Hz).

Τα κύματα βήτα συνδέονται με την ενεργό σκέψη και την εστιασμένη προσοχή. Συχνά κατανέμονται σε μετωπικές και κεντρικές περιοχές. Η αυξημένη δραστηριότητα βήτα μπορεί να σχετίζεται με τον κινητικό σχεδιασμό και αποτελεί επίσης βασικό ρυθμό σε αισθητικοκινητικές εργασίες.

### 5. Γάμμα (γ: 30-70 Hz).

Οι ταλαντώσεις γάμμα μπορεί να αντανακλούν διαδικασίες ολοκλήρωσης, όπως η δέσμευση χαρακτηριστικών και η επιλογή της προσοχής. Αν και περισσότερο συζητημένη, η γάμμα έχει κερδίσει το ενδιαφέρον λόγω του πιθανού ρόλου της στη νόηση υψηλότερου επιπέδου, τη συνειδητή αντίληψη και τον νευρωνικό συγχρονισμό μεταξύ των περιοχών του εγκεφάλου.



Σχήμα 2.2 Οι ζώνες συχνοτήτων σε ΕΕG

Είναι σημαντικό να σημειωθεί ότι αυτά τα όρια είναι κάπως αυθαίρετα- για παράδειγμα, ορισμένοι ερευνητές ορίζουν την άλφα από 8 έως 13 Hz και άλλοι τοποθετούν τις συχνότητες βήτα στα 13 έως 30 Hz. Έχουν προταθεί πρόσθετες ζώνες, όπως η υψηλή γάμμα (70-100+ Hz), οι οποίες μπορούν να διερευνηθούν σε εξειδικευμένες μελέτες ΗΕΓ ή μαγνητοεγκεφαλογραφίας (MEG). Ωστόσο, οι συνήθεις κλινικές καταγραφές ΗΕΓ επικεντρώνονται συνήθως στην περιοχή 0,5-70 Hz, καθώς οι συχνότητες πέραν αυτής της περιοχής συχνά επηρεάζονται από μυϊκά τεχνήματα ή έχουν μειωμένο λόγο σήματος προς θόρυβο.

### Συγχρονισμός ισχύος και φάσης

Εκτός από το περιεχόμενο σθχνότητας, δύο άλλα κρίσιμα χαρακτηριστικά στην ανάλυση του ΗΕΓ είναι η **ισχύς** (που αντικατοπτρίζει το πλάτος ή την ενέργεια του σήματος σε μια δεδομένη ζώνη) και η **φάση** (που δείχνει πώς ευθυγραμμίζονται οι κύκλοι του σήματος σε διαφορετικές φλοιώδεις περιοχές). Ο συγχρονισμός φάσης σε χωρικά διακριτές περιοχές του εγκεφάλου μπορεί να σηματοδοτήσει συνεκτική δραστηριότητα δικτύου και έχει συνδεθεί με διαδικασίες όπως η προσοχή, η μνήμη εργασίας και οι ολοκληρωτικές αισθητηριακές λειτουργίες [8], [9]. Οι μεταβολές σε αυτά τα φασματικά χαρακτηριστικά μπορούν να παρέχουν πληροφορίες για καταστάσεις ασθενειών (π.χ. επιληψία, νόσος του Alzheimer) ή να αντανακλούν αλλαγές που προκαλούνται από εξωτερικά ερεθίσματα (π.χ. ακουστικές, οπτικές ή σωματοαισθητικές εργασίες).

### 2.1.4 Τεχνουργήματα και θόρυβος EEG

Παρά την ευαισθησία και τη μη επεμβατική του φύση, το ΗΕΓ είναι ευαίσθητο σε ένα ευρύ φάσμα **τεχνουργημάτων**- ανεπιθύμητων σημάτων τόσο εξωτερικής όσο και φυσιολογικής προέλευσης [20]. Αυτά τα τεχνουργήματα μπορούν να υποβαθμίσουν την ποιότητα των δεδομένων και να περιπλέξουν την ερμηνεία:

- Περιβαλλοντικές πηγές. Τα ηλεκτρόδια και οι ενισχυτές μπορούν να λαμβάνουν παρεμβολές από γραμμές εναλλασσόμενου ρεύματος (50/60 Hz), από ηλεκτρονικά είδη (υπολογιστές, κινητά τηλέφωνα), ακόμη και από ορισμένα συστήματα φωτισμού. Παρόλο που το σύγχρονο υλικό και οι λύσεις ενεργής θωράκισης βοηθούν στον μετριασμό αυτών των θορύβων, παραμένουν μια διαρκώς παρούσα πρόκληση σε πραγματικές συνθήκες [21].
- 2. Φυσιολογικές πηγές.Πέρα από το φλοιώδες σήμα ενδιαφέροντος, το ανθρώπινο σώμα παράγει πρόσθετη ηλεκτρική δραστηριότητα. Τα τεχνήματαEOG (ηλεκτρο-οφθαλμογραφία) από ανοιγοκλείσιμο των ματιών ή κινήσεις είναι εμφανή στα μετωπιαία ηλεκτρόδια. Μπορεί επίσης να υπάρχουν τεχνουργήματαHKΓ (ηλεκτροκαρδιογραφία) από καρδιακούς παλμούς και σήματα HMΓ (ηλεκτρομυογραφία) από μυϊκή ένταση (π.χ. σφίξιμο των γνάθων, ένταση του αυχένα) [22]. Αυτά τα τεχνουργήματα μπορεί να επικαλύπτονται με τις ίδιες συχνότητες ενδιαφέροντος στις γνωστικές μελέτες EEG.
- 3. Κίνηση και δερματική σύνθετη αντίσταση.Οι κινήσεις των συμμετεχόντων, η μετατόπιση των ηλεκτροδίων και οι διακυμάνσεις στην αγωγιμότητα του τριχωτού της κεφαλής εισάγουν περαιτέρω επιμόλυνση. Ιδιαίτερα σε φορητά σενάρια ΗΕΓ ή σε πραγματικές συνθήκες, τα αντικείμενα κίνησης μπορούν να επισκιάσουν τα υποκείμενα νευρωνικά σήματα εάν δεν ελεγχθούν ή δεν υποβληθούν σε κατάλληλη επεξεργασία.

### 2.1.5 Διόρθωση και προεπεξεργασία τεχνουργημάτων

Για την εξαγωγή ουσιαστικών συμπερασμάτων από τα δεδομένα EEG, οι ερευνητές χρησιμοποιούν μια σειρά από μέτρα, που εκτείνονται από πρωτόκολλα απόκτησης δεδομένων έως μεταγενέστερες υπολογιστικές τεχνικές.

#### Πρωτόκολλα λήψης και προληπτικά μέτρα

 Περιβάλλον καταγραφής. Ιδανικά, το ΗΕΓ καταγράφεται σε ένα θωρακισμένο δωμάτιο (κλωβός Faraday) όπου ελαχιστοποιούνται οι ηλεκτρομαγνητικές παρεμβολές [23]. Ο εξοπλισμός, όπως τα καλώδια, μπορεί να έχει αγώγιμα περιβλήματα και ολόκληρη η διάταξη γειώνεται προσεκτικά για την αποφυγή ηλεκτρικών δυναμικών που δεν σχετίζονται με την εγκεφαλική δραστηριότητα. 2. Προετοιμασία των συμμετεχόντων.Οι συμμετέχοντες έχουν γενικά οδηγίες να μειώσουν την υπερβολική κίνηση, το ανοιγόκλεισμα των ματιών και την ομιλία κατά τη διάρκεια κρίσιμων διαστημάτων εργασίας. Η προετοιμασία του δέρματος (στην περίπτωση των υγρών ηλεκτροδίων) περιλαμβάνει τον καθαρισμό του τριχωτού της κεφαλής με οινόπνευμα ή ένα λειαντικό τζελ για τη μείωση της σύνθετης αντίστασης, βελτιώνοντας την ποιότητα του σήματος.

### Τεχνικές προεπεξεργασίας

- Επαναδειγματοληψία. Αν και οι υψηλοί ρυθμοί δειγματοληψίας αποδίδουν λεπτομερή χρονικά δεδομένα, μπορούν να δημιουργήσουν μεγάλα σύνολα δεδομένων που επιβραδύνουν τα επόμενα βήματα επεξεργασίας. Ως συμβιβασμός, τα δεδομένα ΗΕΓ μπορούν να υποδειγματοληφθούν σε έναν πιο διαχειρίσιμο ρυθμό, διασφαλίζοντας ότι δεν χάνονται κρίσιμες συνιστώσες συχνότητας σύμφωνα με τον κανόνα Nyquist (συχνότητα δειγματοληψίας τουλάχιστον διπλάσια της υψηλότερης συχνότητας ενδιαφέροντος) [24]. Αυτό το βήμα περιλαμβάνει συνήθως ένα φίλτρο αντιδιαστολής για την αποφυγή παραμορφώσεων από υψηλότερες συχνότητες που αναδιπλώνονται πίσω στη ζώνη συχνοτήτων ενδιαφέροντος.
- 2. Επαναφορά. Τα δεδομένα ΗΕΓ πρέπει πάντα να ερμηνεύονται σε σχέση με ένα ηλεκτρόδιο αναφοράς. Οι συνήθεις πρακτικές περιλαμβάνουν την αναφορά σε συνδεδεμένα μαστοειδή (μέσος όρος του σήματος από ηλεκτρόδια που τοποθετούνται πίσω από κάθε αυτί) ή το μέσο όρο όλων των ηλεκτροδίων του τριχωτού της κεφαλής (κοινή μέση αναφορά). Αυτή η επανααναφορά μειώνει τη μεροληψία που εισάγεται από κάθε μεμονωμένη θέση και συχνά μετριάζει τον παγκόσμιο θόρυβο [6].
- 3. Φιλτράρισμα: Το ψηφιακό φιλτράρισμα αποτελεί ακρογωνιαίο λίθο της προεπεξεργασίας του ΗΕΓ. Ο στόχος είναι η εξασθένιση των συχνοτήτων που δεν είναι σαφώς σχετικές με το ερώτημα της μελέτης ή που αντιπροσωπεύουν γνωστές πηγές θορύβου. Στην παρούσα διατριβή χρησιμοποιήθηκαν φίλτρα Chebyshev πεπερασμένης απόκρισης παλμού (FIR) και άπειρης απόκρισης παλμού (IIR) [26].
  - ο Φίλτρο FIR

$$y[n] = \sum_{i=0}^{N} b_i x[n-i]$$
 2.1

Ένα φίλτρο FIR τάξης N υπολογίζει την έξοδο με τη συνέλιξη του σήματος εισόδου με τους συντελεστές b<sub>i</sub> του φίλτρου. Τα φίλτρα FIR είναι εγγενώς σταθερά και γραμμικής φάσης, γεγονός που τα καθιστά δημοφιλή στην επεξεργασία σήματος EEG.

ο Φίλτρο IIR

$$y[n] = \frac{1}{a_0} \left( \sum_{i=0}^{P} b_i x[n-i] - \sum_{j=0}^{Q} a_i y[n-j] \right)$$
(2.2)

όπου P είναι η τάξη του φίλτρου τροφοδότησης, Q είναι η τάξη του φίλτρου ανατροφοδότησης,  $b_i$  είναι οι συντελεστές του φίλτρου τροφοδότησης,  $a_i$  είναι οι συντελεστές του φίλτρου ανατροφοδότησης, x[n] είναι το σήμα εισόδου και y[n] είναι το σήμα εξόδου. Ένα φίλτρο IIR μπορεί να επιτύχει μια καθορισμένη απόκριση φίλτρου με λιγότερους συντελεστές από ένα φίλτρο FIR, αλλά μπορεί να εισάγει παραμορφώσεις φάσης. Με την επιλογή κατάλληλων πόλων και μηδενικών, η παραλλαγή Chebyshev προσφέρει μια πιο απότομη απόκλιση για μια δεδομένη τάξη φίλτρου.

#### ο Συνάρτηση μεταφοράς Chebyshev

$$H(z) = \frac{b_1 + b_2 z^{-1} + \dots + b_{N+1} z^{-N}}{a_1 + a_2 z^{-1} + \dots + a_{N+1} z^{-N}}$$
(2.3)

Χαρακτηρίζεται από μια συμπεριφορά ισοπαλότητας στη ζώνη διέλευσης ή στη ζώνη διακοπής, τα φίλτρα Chebyshev μπορούν να παρέχουν πιο έντονες αποκοπές, αν και πρέπει να δοθεί προσοχή στη διαχείριση των φαινομένων κυμάτωσης.

- 4. Αποστροφή. Γραμμικές ή αργές συνιστώσες ολίσθησης μπορεί να εμφανιστούν στα δεδομένα ΕΕG λόγω μετατοπίσεων του οργάνου, αλλαγών στην αντίσταση των ηλεκτροδίων με την πάροδο του χρόνου ή κινήσεων του συμμετέχοντα. Ο αποπροσανατολισμός αφαιρεί αυτές τις αργές μεταβολές με την προσαρμογή μιας γραμμικής συνάρτησης (συχνά μέσω ελαχίστων τετραγώνων) στα δεδομένα και την αφαίρεσή της [27]. Αυτή η προσέγγιση διασφαλίζει ότι οι επακόλουθες αναλύσεις χρονοσυχνότητας ή ERP δεν στρεβλώνονται από σταδιακές μετατοπίσεις της γραμμής βάσης.
- 5. Ανάλυση ανεξάρτητων συνιστωσών (ICA). Η ICA χρησιμοποιείται ευρέως για τον διαχωρισμό των σημάτων που σχετίζονται με τεχνουργήματα (π.χ. ανοιγοκλείσιμο των ματιών, καρδιακά τεχνουργήματα) από τη γνήσια νευρική δραστηριότητα [28]. Μετά τη λεύκανση του σήματος με την αφαίρεση των συσχετίσεων μεταξύ των συνιστωσών (π.χ. μέσω της διάσπασης ιδιοτιμών του πίνακα συνδιακύμανσης), ο αλγόριθμος ενημερώνει επαναληπτικά το W για να μεγιστοποιήσει την ανεξαρτησία των εξαγόμενων συνιστωσών. Μόλις εντοπιστούν οι συνιστώσες που αντιπροσωπεύουν τεχνουργήματα (π.χ. μέσω συσχέτισής τους με σήματα ΕΟG ή οπτικής επιθεώρησης χαρτών τριχωτού της κεφαλής), μπορούν να αφαιρεθούν, αποδίδοντας ένα «καθαρισμένο» σήμα όταν οι εναπομείνασες συνιστώσες πουβάλλονται πίσω.
- 6. Διόρθωση βασικής γραμμής (baseline correction). Για αναλύσεις που σχετίζονται με γεγονότα, κάθε εποχή συχνά ευθυγραμμίζεται με μια συγκεκριμένη περίοδο βασικής γραμμής πριν από τη διέγερση και η μέση τάση σε αυτή τη βασική γραμμή αφαιρείται από ολόκληρη την εποχή [29]. Με τον τρόπο αυτό, εξαλείφεται οποιαδήποτε μετατόπιση της γραμμής βάσης, διασφαλίζοντας ότι τα επακόλουθα μέτρα πλάτους ή λανθάνουσας κατάστασης αντικατοπτρίζουν τη δυναμική που σχετίζεται με την εργασία και όχι τις αυθαίρετες μετατοπίσεις της τάσης.

Κάθε βήμα προεπεξεργασίας είναι αναπόσπαστο στοιχείο για την επίτευξη δεδομένων ΗΕΓ υψηλής ποιότητας που αποτυπώνουν με ακρίβεια την υποκείμενη ηλεκτρική δραστηριότητα του εγκεφάλου. Η τεχνητή επιμόλυνση μπορεί εύκολα να καλύψει τις λεπτές γνωστικές επιδράσεις ή να παράγει παραπλανητικά αποτελέσματα, ειδικά σε σενάρια που περιλαμβάνουν συνιστώσες υψηλής συχνότητας ή μικρές διαφορές πλάτους. Ο συνδυασμός προσεκτικού πειραματικού ελέγχου (π.χ. μείωση του περιβαλλοντικού θορύβου) και αυστηρών υπολογιστικών μεθόδων (π.χ. φιλτράρισμα, ICA) παρέχει τα θεμέλια για αξιόπιστες αναλύσεις των νευρωνικών διεργασιών.

### 2.1.6 Ανάλυση χρόνου-συχνότητας - Time-Frequency Analysis

Η ανάλυση χρόνου-συχνότητας στο ΗΕΓ είναι μια κρίσιμη και περίπλοκη διαδικασία, απαραίτητη για την αποκάλυψη της πολύπλοκης αλληλεπίδρασης μεταξύ χρόνου και συχνότητας στην εγκεφαλική δραστηριότητα. Η διαδικασία αυτή περιλαμβάνει τη διάσπαση του σήματος ΗΕΓ, το οποίο είναι μια αναπαράσταση της εγκεφαλικής δραστηριότητας στο πεδίο του χρόνου, στις συχνότητες που το συνθέτουν με την πάροδο του χρόνου. Αυτή η διπλή ανάλυση τόσο του πεδίου του χρόνου όσο και της συχνότητας επιτρέπει την ολοκληρωμένη κατανόηση του τρόπου με τον οποίο η εγκεφαλική δραστηριότητα μεταβάλλεται όχι μόνο με την πάροδο του χρόνου αλλά και σε διάφορες ζώνες συχνοτήτων.

Δύο κοινές μέθοδοι που χρησιμοποιούνται στην ανάλυση χρόνου-συχνότητας είναι ο μετασχηματισμός Wavelet και ο μετασχηματισμός Fourier. Ο μετασχηματισμός Wavelet εκτιμάται ιδιαίτερα στην ανάλυση ΗΕΓ για την ικανότητά του να παρέχει μια αναπαράσταση χρόνου-συχνότητας του σήματος ΗΕΓ. Με την αποσύνθεση του σήματος ΗΕΓ σε κυματίδια σε διάφορες συχνότητες και χρονικές στιγμές, προσφέρει μια λεπτομερή εικόνα του τρόπου με τον οποίο οι διάφορες συνιστώσες συχνότητας της ηλεκτρικής δραστηριότητας του εγκεφάλου μεταβάλλονται με την πάροδο του χρόνου. Η μέθοδος αυτή είναι γνωστή για την εξαιρετική χρονική και συχνοτική της ανάλυση, γεγονός που την καθιστά ιδιαίτερα χρήσιμη για την ανάλυση μη στάσιμων σημάτων όπως το ΗΕΓ.Ο μετασχηματισμός Fourier, από την άλλη πλευρά, είναι μια τεχνική που μετατρέπει το σήμα ΗΕΓ στο πεδίο του χρόνου στο πεδίο της συχνότητας. Αυτός ο μετασχηματισμός έχει ως αποτέλεσμα ένα φάσμα που εμφανίζει το εύρος των συχνοτήτων που υπάρχουν στο σήμα ΗΕΓ. Ενώ ο μετασχηματισμός Fourier είναι χρήσιμος για τον εντοπισμό των συνιστωσών συχνότητας του ΗΕΓ, δεν διαθέτει χρονική ανάλυση, γεγονός που την ανάλυση του τρόπου με τον οποίο οι διάφορες ουχιστασχηματισμός έχει ως αποτέλεσμα ένα φάσμα που εμφανίζει το εύρος των συχνοτήτων που υπάρχουν στο σήμα ΗΕΓ. Ενώ ο μετασχηματισμός Fourier είναι χρήσιμος για τον εντοπισμό των συνιστωσών συχνότητας του ΗΕΓ, δεν διαθέτει χρονική ανάλυση, γεγονός που τον καθιστά λιγότερο αποτελεσματικό για την ανάλυση του τρόπου με τον οποίο αυτές οι συχνότητες μεταβάλλονται με την πάροδο του χρόνου.

Στην ανάλυση χρόνου-συχνότητας, η εξαγωγή χαρακτηριστικών όπως η φασματική πυκνότητα ισχύος (Power spectral density PSD) και οι πληροφορίες φάσης είναι κρίσιμες [7]. Η PSD αντιπροσωπεύει την ισχύ κάθε συνιστώσας συχνότητας εντός του σήματος, η οποία είναι θεμελιώδης για τον εντοπισμό των κυρίαρχων συχνοτήτων στο ΗΕΓ και τον τρόπο με τον οποίο αυτές μεταβάλλονται με την πάροδο του χρόνου. Αυτό το χαρακτηριστικό είναι ιδιαίτερα σημαντικό για την κατανόηση της ρυθμικής δραστηριότητας του εγκεφάλου και των διακυμάνσεών της σε διαφορετικές καταστάσεις ή σε απόκριση σε διαφορετικά ερεθίσματα. Η πληροφορία φάσης, η οποία εξετάζει τον χρονισμό των κυματομορφών σε σχέση μεταξύ τους, είναι ζωτικής σημασίας για την κατανόηση του συγχρονισμού και της συνδεσιμότητας μεταξύ διαφορετικών περιοχών του εγκεφάλου. Ο συγχρονισμός φάσης, όπου διαφορετικές περιοχές του εγκεφάλου παρουσιάζουν συγχρονισμένη δραστηριότητα σε συγκεκριμένες συχνότητες, μπορεί να προσφέρει πληροφορίες για το πώς συντονίζονται διαφορετικά τη διάρκεια διαφόρων γνωστικών και αισθητηριακών διεργασιών.
Η συνάφεια των διαφορετικών ζωνών συχνοτήτων στο ΗΕΓ είναι επίσης μια βασική πτυχή της ανάλυσης χρόνου-συχνότητας. Κάθε ζώνη συχνοτήτων, όπως δέλτα (0,5-4 Hz), θήτα (4-8 Hz), άλφα (8-12 Hz), βήτα (12-30 Hz) και γάμμα (30 Hz και άνω), συνδέεται με διαφορετικές εγκεφαλικές καταστάσεις και λειτουργίες. Για παράδειγμα, τα κύματα δέλτα συνδέονται συχνά με τον βαθύ ύπνο, ενώ τα κύματα άλφα παρατηρούνται συνήθως κατά τη διάρκεια της χαλάρωσης. Η κατανόηση αυτών των ζωνών και η σημασία τους για τις διάφορες εγκεφαλικές λειτουργίες είναι ζωτικής σημασίας για την ερμηνεία των δεδομένων ΗΕΓ.

Στη παρόυσα διατριβή έχουμε δώσει ιδιαίτερη σημασία και έχει χρησιμοποιηθεί η ανάλυση με τη χρήση των Morlet Wavelets. Τα Morlet wavelets είναι ένα ημιτονοειδές σήμα εγκλεισμένο σε ένα Gausian παράθυρο (Gaussian envelope), όπως φαίνεται στο Σχήμα 2.3, και ακολουθούν τη διατύπωση των Tallon-Baudry et al. [8]. Στη φασματοχρονική ανάλυση (όπως και με τις παραδοσιακές μεθόδους Fourier), η χρονική και η φασματική ανάλυση είναι αλληλένδετες: μεγαλύτερα χρονικά παράθυρα επιτρέπουν ακριβέστερες εκτιμήσεις συχνότητας· μικρότερα χρονικά παράθυρα «θολώνουν» (smear) τις εκτιμήσεις συχνότητας, παρέχοντας παράλληλα ακριβέστερες χρονικές πληροφορίες. Οι χρονο-συχνοτικές αναπαραστάσεις υπολογίζονται χρησιμοποιώντας ένα κινούμενο χρονικό παράθυρο (sliding window).



Είτε το χρονικό παράθυρο έχει σταθερό μήκος ανεξάρτητα από τη συχνότητα, είτε το χρονικό παράθυρο μειώνεται σε μήκος με την αύξηση της συχνότητας.



Σχήμα 2.4 Χρονική και συχνοτική εξομάλυνση. (α) Για ένα χρονικό παράθυρο σταθερού μήκους, η χρονική και η συχνοτική εξομάλυνση παραμένουν σταθερές. (β) Για χρονικά παράθυρα που μειώνονται με τη συχνότητα, η χρονική εξομάλυνση μειώνεται και η συχνοτική εξομάλυνση αυξάνεται με τη συχνότητα. Πηγή <u>Fieltrip</u>

Mε τη χρήση του προγγραμματιστικού εργαλείου που επιλέχθηκε, το MNE-Python, το μήκος του κυματιδίου Morlet επηρεάζεται από τα ορίσματα *freqs* και *n\_cycles*, τα οποία καθορίζουν τις συχνότητες ενδιαφέροντος και τον αριθμό των κύκλων, αντίστοιχα. Για τη χρονοσυχνοτική αναπαράσταση, το μήκος του κυματιδίου ορίζεται έτσι ώστε και οι δύο ουρές (tails) του κυματιδίου να εκτείνονται κατά πέντε τυπικές αποκλίσεις από το μέσο σημείο της Γκαουσιανής του περιβάλλουσας (Gaussian envelope) και να υπάρχει ένα δείγμα στον χρόνο μηδέν. Το μήκος του κυματιδίου είναι επομένως 10 × sfreq · σ - 1 το οποίο ισούται με  $\frac{5}{\pi} \cdot \frac{n_ccycles \cdot sfreq}{freqs} - 1$  όπου  $\sigma = \frac{n_ccycles}{2\pi f}$  αντιστοιχεί στην τυπική απόκλιση του Gaussian envelope του κυματιδίου. Για παράδειγμα, Η ελάττωση της τιμής *n\_cycles* θα αυξήσει τη χρονική ανάλυση εις βάρος της συχνοτικής ανάλυσης, και αντιστρόφως. Το φασματικό εύρος ζώνης είναι 30/72 = 8,6 Hz), ενώ η διάρκεια του κυματιδίου ισούται με *width/F/π* (σε αυτή την περίπτωση, 7/30/π = 0,074s = 74ms)

Συνοψίζοντας, η ανάλυση χρονοσυχνοτήτων στο ΗΕΓ είναι μια πολύπλοκη αλλά ουσιαστική διαδικασία που παρέχει μια πολυδιάστατη εικόνα της εγκεφαλικής δραστηριότητας. Με την ανάλυση τόσο του χρονικού όσο και του συχνοτικού περιεχομένου της εγκεφαλικής δραστηριότητας, οι ερευνητές μπορούν να αποκτήσουν βαθύτερη κατανόηση των δυναμικών διεργασιών του εγκεφάλου. Αυτή η μεθοδολογική προσέγγιση είναι θεμελιώδης για την προώθηση των γνώσεών μας σχετικά με τη λειτουργία του εγκεφάλου και τις αντιδράσεις του σε διάφορα ερεθίσματα, καθιστώντας την ανεκτίμητο εργαλείο τόσο σε ερευνητικό όσο και σε κλινικό επίπεδο.

#### 2.1.7 Απεικόνιση ηλεκτρικής πηγής - EEG Source Localization

Η απεικόνιση ηλεκτρικής πηγής (EEG Source Localization - ESL), μια βασική τεχνική στην ανάλυση του ΗΕΓ, αποτελεί σημαντικό άλμα στον τομέα των νευροεπιστημών, ιδίως στον τομέα της ερμηνείας και του εντοπισμού της εγκεφαλικής δραστηριότητας. Η ESL ουσιαστικά γεφυρώνει το χάσμα μεταξύ των ηλεκτρικών σημάτων που ανιχνεύονται στο τριχωτό της κεφαλής και της πραγματικής προέλευσής τους εντός της περίπλοκης δομής του εγκεφάλου. Η τεχνική αυτή βασίζεται στην ενσωμάτωση τόσο των χρονικών όσο και των χωρικών πτυχών των σημάτων ΗΕΓ, επιτρέποντας μια ολοκληρωμένη ανάλυση που υπερβαίνει τις δυνατότητες άλλων μεθόδων νευροαπεικόνισης, όπως η fMRI. Η αρχή του εντοπισμού της πηγής στην ανάλυση του ΗΕΓ χρονολογείται από τις απαρχές του ίδιου του ΗΕΓ.

Μία από τις κεντρικές προκλήσεις στον εντοπισμό πηγών ΗΕΓ είναι το αντίστροφο πρόβλημα - η εξαγωγή συμπερασμάτων για την προέλευση της εγκεφαλικής δραστηριότητας από τα σήματα ΗΕΓ που καταγράφονται στο τριχωτό της κεφαλής [9]. Το πρόβλημα αυτό είναι εγγενώς ill-posed λόγω του δυσανάλογου αριθμού των ηλεκτροδίων που χρησιμοποιούνται στις καταγραφές ΗΕΓ ως προς τον τεράστιο αριθμό των νευρώνων στον εγκέφαλο. Κάθε ηλεκτρόδιο κατγράφει τα αθροιστικά ηλεκτρικά δυναμικά που παράγονται από σύγχρονες/ταυτόχρονες νευρωνικές δραστηριότητες, τα οποία στη συνέχεια διαδίδονται μέσω διαφόρων ιστών πριν καταγραφούν. Για την επίλυση αυτού του πολύπλοκου αντίστροφου προβλήματος χρησιμοποιούνται μαθηματικές διατυπώσεις όπως η εξίσωση του Poisson.

$$\nabla^2 \varphi = -\frac{\rho_f}{\varepsilon} \tag{2.4}$$

Αυτή η εξίσωση, που παράγεται από τις εξισώσεις του Maxwell, υποθέτει ότι οι ιστοί της κεφαλής διαθέτουν τη διαπερατότητα του ελεύθερου χώρου. Η λύση της εξίσωσης βασίζεται σε μεγάλο βαθμό στην ακριβή αναπαράσταση των ηλεκτρικών και γεωμετρικών ιδιοτήτων του κεφαλιού - παράγοντες που είναι κρίσιμοι για την κατασκευή ενός ακριβούς μοντέλου κεφαλιού. Αυτά τα μοντέλα συχνά αξιοποιούν δεδομένα μαγνητικής τομογραφίας για την απόκτηση δεδομένων μαγνητικής τομογραφίας ειδικά για κάθε άτομο, με αποτέλεσμα οι ερευνητές να βασίζονται συχνά σε τυποποιημένα ανατομικά πρότυπα.

Ένα θεμελιώδες ζήτημα στον εντοπισμό πηγών ΕΕG είναι η μη μοναδικότητα των λύσεων. Τα ίδια σήματα ΗΕΓ που έχουν καταγραφεί από το τριχωτό της κεφαλής μπορούν να παραχθούν από ένα τεράστιο φάσμα συνδυασμών νευρικών πηγών, ακόμη και υπό ιδανικές συνθήκες με τέλεια μοντέλα κεφαλής και αθόρυβα σήματα [10]. Για να το αντιμετωπίσουν αυτό, οι ερευνητές επιβάλλουν μαθηματικούς, ανατομικούς ή νευροφυσικούς περιορισμούς στις πιθανές θέσεις και προσανατολισμούς των νευρικών πηγών, μειώνοντας έτσι την υπολογιστική πολυπλοκότητα και ενισχύοντας την ακρίβεια της λύσης. Οι πρόσφατες εξελίξεις στον τομέα, ιδίως η ενσωμάτωση τεχνικών μηχανικής μάθησης, έχουν φέρει επανάσταση στην προσέγγιση της επίλυσης του αντίστροφου προβλήματος. Αυτές οι νέες μεθοδολογίες έχουν μειώσει σημαντικά το χρόνο και τους υπολογιστικούς πόρους που απαιτούνται για τον ακριβή εντοπισμό των πηγών. Οι εξελιγμένοι αλγόριθμοι ελαχιστοποιούν πλέον αποτελεσματικά τα σφάλματα εντοπισμού, επιτρέποντας ακριβέστερες εκτιμήσεις των ενεργοποιήσεων των φλοιωδών περιοχών. Η πρόοδος αυτή είναι εμφανής σε ένα φάσμα εφαρμογών, από κλινικές χρήσεις στη θεραπεία της επίληψίας έως γνωστικές μελέτες που αποσκοπούν στην κατανόηση των συναισθηματικών αντιδράσεων.

#### Ευθύ και Αντίστροφο πρόβλημα

#### Ευθύ Πρόβλημα

Το ευθύ πρόβλημα ορίζεται ως ο υπολογισμός των ηλεκτρικών δυναμικών στην επιφάνεια του τριχωτού της κεφαλής (ή των μαγνητικών πεδίων εκτός αυτής στην περίπτωση της μαγνητοεγκεφαλογραφίας - ΜΕΓ), δεδομένης μιας συγκεκριμένης κατανομής πηγών ηλεκτρικής δραστηριότητας εντός του εγκεφάλου και ενός μοντέλου της κεφαλής ως αγωγού όγκου (volume conductor) [Mosher, J. C., Leahy, R. M., & Lewis, P. S. (1999). EEG and MEG: forward solutions for inverse methods. IEEE Transactions on Biomedical Engineering, 46(3), 245-259]. Η ακρίβεια της λύσης του ευθέος προβλήματος είναι υψίστης σημασίας, καθώς οποιαδήποτε ανακρίβεια ή σφάλμα στο στάδιο αυτό θα πολλαπλασιαστεί και θα επηρεάσει δραματικά την αξιοπιστία των αποτελεσμάτων του αντίστροφου προβλήματος, δηλαδή την εκτίμηση της θέσης και της ισχύος των εγκεφαλικών πηγών.

Η φυσική που διέπει τη διάδοση των ηλεκτρικών πεδίων μέσω των ιστών της κεφαλής, υπό την παραδοχή των οιονεί στατικών συνθηκών (quasi-static approximation) για τις βιολογικές συχνότητες του ΗΕΓ, περιγράφεται από την εξίσωση Poisson, η οποία προκύπτει από τις εξισώσεις Maxwell:

$$\nabla \cdot (\sigma(r) \nabla \Phi(r)) = -l_{\nu}(r) \tag{2.5}$$

όπου:

- σ(r) είναι η τανυστής ηλεκτρικής αγωγιμότητας (conductivity tensor) στη θέση r. Για ισοτροπικούς ιστούς, απλοποιείται σε μια βαθμωτή τιμή αγωγιμότητας.
- *l<sub>v</sub>(r)* είναι η πυκνότητα του πρωτεύοντος ρεύματος (primary current density) που παράγεται από τις νευρωνικές πηγές στη θέση r.
- ∇ · είναι ο τελεστής απόκλισης (divergence operator) και ∇ είναι ο τελεστής κλίσης (gradient operator).

Η επίλυση αυτής της εξίσωσης απαιτεί ακριβή γνώση των γεωμετρικών και ηλεκτρικών ιδιοτήτων των διαφόρων ιστών της κεφαλής.

#### Μοντελοποίηση Κεφαλής (Head Modeling)

Για την επίλυση του ευθέος προβλήματος, είναι απαραίτητη η κατασκευή ενός ρεαλιστικού μοντέλου κεφαλής που να αναπαριστά τη γεωμετρία και τις ηλεκτρικές ιδιότητες των διαφόρων ιστών που διαπερνά το ηλεκτρικό σήμα από τις νευρωνικές πηγές έως τα επιφανειακά ηλεκτρόδια. Οι κυριότεροι ιστοί που λαμβάνονται υπόψη είναι το τριχωτό της κεφαλής (scalp), το κρανίο (skull), το εγκεφαλονωτιαίο υγρό (ENY - cerebrospinal fluid, CSF) και ο εγκεφαλικός ιστός (brain).

Στη βιβλιογραφία συναντώνται δύο κύριες κατηγορίες μοντέλων κεφαλής. Η πρώτη κατηγορία περιλαμβάνει τα Σφαιρικά Movτέλα (Spherical Models), τα οποία αποτελούν απλοποιημένες αναπαραστάσεις όπου η κεφαλή προσεγγίζεται από μία ή περισσότερες ομόκεντρες σφαίρες, καθεμία εκ των οποίων αντιστοιχεί σε έναν τύπο ιστού με ομοιογενή και ισοτροπική αγωγιμότητα [11]. Παρότι τα σφαιρικά μοντέλα είναι υπολογιστικά αποδοτικά, εισάγουν σημαντικές ανακρίβειες, ιδιαίτερα σε περιοχές μακριά από το κέντρο της κεφαλής ή σε άτομα με μη τυπική γεωμετρία κρανίου.

Η δεύτερη κατηγορία αφορά τα Ρεαλιστικά Μοντέλα (Realistic Models), τα οποία βασίζονται σε ανατομικές εικόνες του ατόμου, συνήθως από Μαγνητική Τομογραφία (ΜΤ -Magnetic Resonance Imaging, MRI). Οι εικόνες αυτές υποβάλλονται σε διαδικασία τμηματοποίησης (segmentation) για τον ακριβή προσδιορισμό των ορίων μεταξύ των διαφόρων ιστών. Για την επίλυση της εξίσωσης Poisson σε αυτές τις ρεαλιστικές γεωμετρίες, οι πιο διαδεδομένες αριθμητικές μέθοδοι είναι η Μέθοδος Οριακών Στοιχείων (Boundary Element Method - BEM) και η Μέθοδος Πεπερασμένων Στοιχείων (Finite Element Method - FEM). Η Μέθοδος Οριακών Στοιγείων (BEM) απαιτεί την τμηματοποίηση μόνο των επιφανειών που διαγωρίζουν περιογές με διαφορετική, αλλά εντός κάθε περιογής ομοιογενή, αγωγιμότητα. Καθίσταται έτσι ιδιαίτερα κατάλληλη για μοντέλα με λίγα ομοιογενή διαμερίσματα [12]. Από την άλλη πλευρά, η Μέθοδος Πεπερασμένων Στοιχείων (FEM) επιτρέπει τη μοντελοποίηση πιο πολύπλοκων γεωμετριών και την ενσωμάτωση ανισότροπων αγωγιμοτήτων, όπου η αγωγιμότητα διαφέρει ανάλογα με την κατεύθυνση, όπως παρατηρείται για παράδειγμα στον λευκό εγκεφαλικό ιστό. Παρότι είναι υπολογιστικά πιο απαιτητική, η FEM είναι δυνητικά πιο ακριβής στην αναπαράσταση των πολύπλοκων ηλεκτρικών ιδιοτήτων των ιστών [13].



Σχήμα 2.5 Head model με σωστη τοποθέτηση ηλεκτροδίων (κόκκινα σημεία). Η τμηματοποίηση και η δημιουργία του μοντέλου έγινε με χρήση της μεθόδου BEM

Ανεξαρτήτως της επιλεγόμενης μεθόδου μοντελοποίησης, η ακρίβεια των τιμών αγωγιμότητας που αποδίδονται σε κάθε ιστό είναι κρίσιμη. Αυτές οι τιμές συχνά λαμβάνονται από τη βιβλιογραφία, αλλά παρουσιάζουν σημαντική διακύμανση μεταξύ ατόμων και μεθόδων μέτρησης, αποτελώντας έτσι μια σημαντική πηγή αβεβαιότητας στην ακρίβεια του ευθέος προβλήματος [14].

#### Ο Πίνακας Πεδίου Αγωγιμότητας (Leadfield Matrix)

Αφού οριστεί το μοντέλο των πηγών (συνήθως ως ισοδύναμα ρεύματα διπόλων - equivalent current dipoles, ECDs) και το μοντέλο αγωγού όγκου της κεφαλής, ο επόμενος στόχος του ευθέος προβλήματος είναι ο υπολογισμός του πώς η δραστηριότητα κάθε πιθανής πηγής εντός του εγκεφάλου συνεισφέρει στα δυναμικά που μετρώνται από κάθε ηλεκτρόδιο ΗΕΓ. Αυτή η θεμελιώδης σχέση ποσοτικοποιείται μέσω του Πίνακα Πεδίου Αγωγιμότητας (Leadfield Matrix - LFM), ο οποίος συχνά συμβολίζεται στη βιβλιογραφία με L ή G.

Για ένα σύνολο  $N_e$  ηλεκτροδίων και  $N_s$  πιθανών θέσεων πηγών (ή προσανατολισμών διπόλων, στην περίπτωση που λαμβάνεται υπόψη και ο προσανατολισμός της πηγής), ο LFM είναι ένας πίνακας διαστάσεων  $N_e \times N_s$ . Κάθε στήλη του LFM αντιστοιχεί σε μία μοναδική πηγή και περιγράφει τα ηλεκτρικά δυναμικά που θα παρήγαγε αυτή η πηγή, υποθέτοντας μοναδιαία ισχύ, σε όλα τα καταγραφικά ηλεκτρόδια. Η σχέση μεταξύ των μετρούμενων δυναμικών V (ένα διάνυσμα διαστάσεων  $N_e \times 1$ ), του Πίνακα Πεδίου Αγωγιμότητας L, και των στιγμών των πηγών J (ένα διάνυσμα διαστάσεων  $N_s \times 1$  που αντιπροσωπεύει την ισχύ και τον προσανατολισμό των πηγών) μπορεί να εκφραστεί με τη γραμμική εξίσωση:

$$\mathbf{V} = \mathbf{L} \mathbf{J} + \boldsymbol{\varepsilon} \tag{2.6}$$

όπου ο όρος ε αντιπροσωπεύει τον θόρυβο μέτρησης ή τα σφάλματα που προκύπτουν από τις προσεγγίσεις της μοντελοποίησης.

Ο ακριβής υπολογισμός του LFM είναι μια πολύπλοκη διαδικασία που εξαρτάται από την ακριβή γνώση αρκετών παραμέτρων. Πρωτίστως, απαιτείται η λεπτομερής γεωμετρία του μοντέλου κεφαλής, δηλαδή οι ακριβείς συντεταγμένες των ορίων των διαφόρων ιστών. Δεύτερον, οι τιμές ηλεκτρικής αγωγιμότητας που αποδίδονται σε κάθε διαμέρισμα ιστού είναι κρίσιμες για τον καθορισμό της διάδοσης του ρεύματος. Τρίτον, οι ακριβείς τρισδιάστατες συντεταγμένες των ηλεκτροδίων ΗΕΓ στην επιφάνεια του τριχωτού της κεφαλής, είτε λαμβανόμενες με εξειδικευμένα συστήματα ψηφιοποίησης (digitizers) είτε προσεγγιζόμενες από πρότυπα συστήματα τοποθέτησης ηλεκτροδίων (όπως το διεθνές σύστημα 10-20), πρέπει να είναι διαθέσιμες και συν-καταγεγραμμένες με το μοντέλο κεφαλής. Τέλος, ο ορισμός του χώρου πηγών (source space), δηλαδή του συνόλου των πιθανών θέσεων και, ενδεχομένως, των προσανατολισμών των νευρωνικών πηγών εντός του εγκεφαλικού όγκου, ολοκληρώνει τις απαραίτητες πληροφορίες για τον υπολογισμό του LFM.

Η ακρίβεια του Πίνακα Πεδίου Αγωγιμότητας είναι θεμελιώδους σημασίας, καθώς οποιαδήποτε σφάλματα στη γεωμετρία του μοντέλου, στις παραδοχές για τις τιμές αγωγιμότητας, ή στην εκτίμηση της τοποθέτησης των ηλεκτροδίων θα οδηγήσουν αναπόφευκτα σε έναν ανακριβή LFM. Ένας τέτοιος ανακριβής πίνακας, με τη σειρά του, θα υπονομεύσει σημαντικά την ποιότητα και την αξιοπιστία των λύσεων που θα προκύψουν από το αντίστροφο πρόβλημα, επηρεάζοντας άμεσα την ικανότητα ακριβούς εντοπισμού των εγκεφαλικών πηγών [15].

#### Μέθοδος Εφαρμογής Διπόλου

Η προσέγγιση αυτή βασίζεται στην αντίληψη ότι η ηλεκτρική δραστηριότητα του εγκεφάλου παράγεται από ένα ή περισσότερα δίπολα - υποθετικά μεμονωμένα σημεία νευρικής δραστηριότητας. Κάθε δίπολο σε αυτό το μοντέλο χαρακτηρίζεται από παραμέτρους όπως η θέση του εντός του εγκεφάλου, ο προσανατολισμός του και η ισχύς ή το μέγεθος της δραστηριότητάς του. Η πρωταρχική λειτουργία της μοντελοποίησης διπόλων είναι η προσαρμογή αυτών των υποθετικών διπόλων στα δεδομένα ΗΕΓ, συμπεραίνοντας έτσι τις πιθανές πηγές της καταγεγραμμένης εγκεφαλικής δραστηριότητας [16]. Η μέθοδος αυτή λειτουργεί με την υπόθεση ότι διακριτές περιοχές του εγκεφάλου είναι υπεύθυνες για τη δημιουργία των παρατηρούμενων σημάτων ΗΕΓ. Εντοπίζοντας αυτές τις περιοχές, η μοντελοποίηση διπόλων παρέχει πολύτιμες πληροφορίες για τις χωρικές πτυχές της λειτουργίας του εγκεφάλου. Ωστόσο, η μοντελοποίηση διπόλου δεν είναι χωρίς περιορισμούς. Μία από τις κύριες επικρίσεις αυτής της προσέγγισης είναι η πιθανότητα υπεραπλούστευσης της πολύπλοκης ηλεκτρικής δραστηριότητας του εγκεφάλου. Ο εγκέφαλος είναι ένα εξαιρετικά διασυνδεδεμένο όργανο, με νευρωνικά δίκτυα που εξαπλώνονται σε διάφορες περιοχές. Με τη μείωση αυτών των δικτύων σε λίγα διακριτά σημεία, τα μοντέλα διπόλου μπορεί να μην αποτυπώνουν με ακρίβεια την κατανεμημένη φύση της εγκεφαλικής δραστηριότητας. Αυτή η απλούστευση μπορεί να οδηγήσει σε απώλεια πληροφοριών, ιδίως στη μελέτη πολύπλοκων γνωστικών διαδικασιών που περιλαμβάνουν πολλαπλές περιοχές του εγκεφάλου [17]. Επιπλέον, η ακρίβεια της μοντελοποίησης διπόλων εξαρτάται σε μεγάλο βαθμό από τη σωστή αναγνώριση και τον χαρακτηρισμό των ίδιων των διπόλων. Αυτό περιλαμβάνει ακριβείς εκτιμήσεις των θέσεων, των προσανατολισμών και των δυνάμεών τους, κάτι που μπορεί να αποτελέσει πρόκληση δεδομένης της πολυπλοκότητας της ανατομίας του εγκεφάλου και της μεταβλητότητας της ποιότητας του σήματος ΗΕΓ.

Παρά τις προκλήσεις αυτές, η μοντελοποίηση διπόλων παραμένει μια θεμελιώδης προσέγγιση στον εντοπισμό πηγών ΗΕΓ. Η ευθύτητά της και η αποτελεσματικότητά της σε ορισμένα σενάρια εξακολουθούν να την καθιστούν πολύτιμη μέθοδο στον τομέα. Ωστόσο, οι ερευνητές και οι κλινικοί γιατροί πρέπει να γνωρίζουν τους περιορισμούς της, ιδίως όταν ερμηνεύουν τα αποτελέσματα στο πλαίσιο πιο σύνθετων δυναμικών του εγκεφάλου. Καθώς η νευροεπιστήμη συνεχίζει να εξελίσσεται, η μοντελοποίηση του διπόλου συμπληρώνεται και μερικές φορές αντικαθίσταται από πιο εξελιγμένες τεχνικές που μπορούν να αποτυπώσουν την κατανεμημένη φύση της εγκεφαλικής δραστηριότητας. Παρ' όλα αυτά, ο ρόλος της στην ιστορική και συνεχιζόμενη ανάπτυξη των τεχνικών εντοπισμού των πηγών του ΗΕΓ υπογραμμίζει τη σημασία της στον τομέα.

#### Κατανεμημένες Μέθοδοι

Τα μοντέλα κατανεμημένων πηγών αντιπροσωπεύουν μια αλλαγή παραδείγματος στον εντοπισμό πηγών ΗΕΓ, σε έντονη αντίθεση με την έννοια της μοντελοποίησης διπόλου. Ενώ η μοντελοποίηση διπόλου επικεντρώνεται στον εντοπισμό διακριτών πηγών εγκεφαλικής δραστηριότητας, τα μοντέλα κατανεμημένης πηγής προσεγγίζουν την εγκεφαλική δραστηριότητα ως συνεχές πεδίο σε ολόκληρο τον εγκέφαλο ή σε σημαντικές περιοχές του. Αυτή η προοπτική επιτρέπει μια πιο ολιστική αναπαράσταση της εγκεφαλικής δραστηριότητας, αναγνωρίζοντας την πολυπλοκότητα και την κατανεμημένη φύση των νευρωνικών διεργασιών.

Οι βασικές τεχνικές στο πλαίσιο των μοντέλων κατανεμημένης πηγής περιλαμβάνουν την ηλεκτρομαγνητική τομογραφία εγκεφάλου χαμηλής ανάλυσης (LORETA) [18] και τις προηγμένες παραλλαγές της, όπως η τυποποιημένη LORETA (sLORETA) [19]και η ακριβής LORETA (eLORETA) [20]. Αυτές οι μέθοδοι έχουν σχεδιαστεί για να παρέχουν μια χωρικά εκτεταμένη άποψη της εγκεφαλικής δραστηριότητας, αποτυπώνοντας τις αποχρώσεις τόσο των εστιακών όσο και των κατανεμημένων νευρικών πηγών. Η μέθοδος LORETA και οι παραλλαγές της λειτουργούν με τον υπολογισμό της τρισδιάστατης κατανομής της ηλεκτρικής δραστηριότητας στον εγκέφαλο. Χρησιμοποιούν αλγορίθμους που εκτιμούν τις νευρικές πηγές των σημάτων ΗΕΓ που καταγράφονται στο τριχωτό της κεφαλής, προβάλλοντας τα σήματα αυτά πίσω στον εγκέφαλο για τον εντοπισμό πιθανών περιοχών δραστηριότητας. Η διαδικασία αυτή περιλαμβάνει περίπλοκους μαθηματικούς υπολογισμούς που λαμβάνουν υπόψη τις αγώγιμες ιδιότητες των διαφόρων ιστών του κεφαλιού και την πολύπλοκη γεωμετρική σχέση μεταξύ του εγκεφάλου και των ηλεκτροδίων του τριχωτού της κεφαλής.

Ένα από τα σημαντικά πλεονεκτήματα των μοντέλων κατανεμημένης πηγής είναι η ικανότητά τους να παρέχουν μια ολοκληρωμένη εικόνα της εγκεφαλικής δραστηριότητας. Αυτό το χαρακτηριστικό είναι ιδιαίτερα πολύτιμο σε μελέτες που διερευνούν σύνθετα γνωστικά καθήκοντα που εμπλέκουν ταυτόχρονα πολλές περιοχές του εγκεφάλου ή κατά τη διερεύνηση εκτεταμένων νευρωνικών δικτύων. Ωστόσο, η εφαρμογή των κατανεμημένων μοντέλων πηγής δεν είναι χωρίς προκλήσεις. Τα μοντέλα αυτά είναι υπολογιστικά εντατικά, απαιτώντας σημαντική επεξεργαστική ισχύ και εξελιγμένους αλγορίθμους για την ακριβή εκτίμηση του ευρέος φάσματος πιθανών νευρωνικών πηγών. Για τη διαγείριση αυτών των προκλήσεων, στα μοντέλα κατανεμημένων πηγών χρησιμοποιούνται συχνά τεχνικές κανονικοποίησης. Η κανονικοποίηση περιλαμβάνει την εφαρμογή περιορισμών ή πρόσθετων πληροφοριών στα μοντέλα, ώστε η λύση του αντίστροφου προβλήματος να είναι πιο σταθερή και αξιόπιστη. Οι τεχνικές αυτές βοηθούν στον μετριασμό του προβλήματος της μη μοναδικότητας στις λύσεις, διασφαλίζοντας ότι οι εκτιμώμενες κατανομές πηγών είναι φυσικά και βιολογικά αληθοφανείς. Ενώ τα μοντέλα κατανεμημένων πηγών προσφέρουν μια πιο λεπτομερή και εκτεταμένη άποψη της εγκεφαλικής δραστηριότητας, η ισορροπία μεταξύ υπολογιστικής έντασης και ακρίβειας αποτελεί κρίσιμο ζήτημα.



Σχήμα 2.6 Παράδειγμα εντοπισμού πηγής με χρήση της μεθόδου sLORETA

#### **Beamformers**

Η μέθοδος beamformers αποτελεί σημαντική πρόοδο στον τομέα του εντοπισμού πηγών ΗΕΓ, προσφέροντας μια εξελιγμένη προσέγγιση για τη χαρτογράφηση της εγκεφαλικής δραστηριότητας. Η τεχνική αυτή χρησιμοποιεί αλγορίθμους χωρικού φιλτραρίσματος για τον ακριβή εντοπισμό των πηγών των σημάτων που καταγράφονται από το ΗΕΓ, μια διαδικασία που βελτιώνει σημαντικά την κατανόηση τόσο των συγγρονισμένων όσο και των ασύγχρονων νευρωνικών δραστηριοτήτων [21]. Στον πυρήνα της, η διαμόρφωση δέσμης βασίζεται στη χρήση μαθηματικών αλγορίθμων για την επεξεργασία των χωρικών πληροφοριών που περιέχονται στα σήματα ΗΕΓ. Με την εφαρμογή αυτών των αλγορίθμων, η διαμόρφωση δέσμης μπορεί να απομονώσει και να εντοπίσει αποτελεσματικά τα νευρωνικά σήματα που προέρχονται από διαφορετικές περιοχές του εγκεφάλου. Αυτή η τεχνική είναι ικανή να παρέχει χωρική χαρτογράφηση υψηλής ανάλυσης της εγκεφαλικής δραστηριότητας, καθιστώντας την ιδιαίτερα χρήσιμη για μελέτες που απαιτούν λεπτομερή εντοπισμό δυναμικών εγκεφαλικών διεργασιών. Η εφαρμογή της διαμόρφωσης δέσμης περιλαμβάνει τη δημιουργία χωρικών φίλτρων που μπορούν να εστιάζουν σε συγκεκριμένες περιοχές του εγκεφάλου, όπως ένα κατευθυντικό μικρόφωνο που απομονώνει τους ήχους από μια συγκεκριμένη κατεύθυνση. Αυτά τα φίλτρα αναλύουν τα δεδομένα ΗΕΓ, διαγωρίζοντας το σήμα ενδιαφέροντος (που προέργεται από μια συγκεκριμένη περιογή του εγκεφάλου) από άλλες άσχετες εγκεφαλικές δραστηριότητες και θόρυβο. Αυτή η επιλεκτική απομόνωση είναι που καθιστά τη διαμόρφωση δέσμης ιδιαίτερα ισχυρή στον εντοπισμό της εγκεφαλικής δραστηριότητας με υψηλό βαθμό ακρίβειας.

Ο Linearly Constrained Minimum Variance (LCMV) αλγόριθμος ξεχωρίζει ως μια ιδιαίτερα προηγμένη και αποτελεσματική τεχνική στο πεδίο της διαμόρφωσης δέσμης για τον εντοπισμό πηγής ΗΕΓ. Η μέθοδος αυτή αποτελεί παράδειγμα της αιχμής της τεχνολογίας χωρικού φιλτραρίσματος, προσφέροντας ακριβείς και δυναμικές γνώσεις σχετικά με την εγκεφαλική δραστηριότητα. Ο διαμορφωτής δέσμης LCMV έχει σχεδιαστεί ειδικά για να απομονώνει νευρικά σήματα από συγκεκριμένες περιοχές του εγκεφάλου, ενώ παράλληλα καταστέλλει ενεργά το θόρυβο και τα σήματα από άλλες περιοχές, παρέχοντας έτσι μια σαφή και εστιασμένη εικόνα της εγκεφαλικής δραστηριότητας.

Ο διαμορφωτής δέσμης LCMV λειτουργεί με τη δημιουργία χωρικών φίλτρων που περιορίζονται γραμμικά ώστε να εστιάζουν σε μια πηγή-στόχο, ελαχιστοποιώντας παράλληλα τη συμβολή άλλων πηγών. Αυτό επιτυγχάνεται μέσω μιας διαδικασίας βελτιστοποίησης όπου ελαχιστοποιείται η διακύμανση του σήματος εξόδου, υπό τον περιορισμό ότι το κέρδος στην κατεύθυνση της πηγής-στόχου παραμένει σταθερό. Αυτή η προσέγγιση επιτρέπει στον διαμορφωτή δέσμης LCMV να απομονώνει αποτελεσματικά τη δραστηριότητα ενδιαφέροντος από την πληθώρα σημάτων που καταγράφονται από τα ηλεκτρόδια EEG.

Η βασική εξίσωση του διαμορφωτή δέσμης LCMV, η οποία χρησιμοποιείται για τον υπολογισμό του χωρικού φίλτρου, είναι απαραίτητη. Μπορεί να αναπαρασταθεί ως εξής:

$$w = S^{-1}C(C^{H}S^{-1}C)^{-1}R$$
 2.7

όπου w είναι το διάνυσμα βάρους του διαμορφωτή δέσμης, R είναι ο πίνακας συνδιακύμανσης των δεδομένων EEG, και S είναι το διάνυσμα lead field για τη θέση της πηγής-στόχου.

Ένα από τα βασικά πλεονεκτήματα της τεχνικής του σχηματιστή δέσμης LCMV στον εντοπισμό πηγής EEG είναι η ευρωστία της στην αντιμετώπιση πολύπλοκων νευρωνικών δραστηριοτήτων. Σε αντίθεση με ορισμένες άλλες μεθόδους διαμόρφωσης δέσμης, ο διαμορφωτής δέσμης LCMV μπορεί να γειριστεί καταστάσεις όπου υπάρχουν πολλαπλές ενεργές πηγές ή όπου ο λόγος σήματος προς θόρυβο είναι χαμηλός. Αυτό τον καθιστά ιδιαίτερα χρήσιμο σε κλινικά και ερευνητικά σενάρια όπου τα νευρωνικά σήματα ενδιαφέροντος είναι ανεπαίσθητα ή ενσωματωμένα σε ένα υπόβαθρο άλλων εγκεφαλικών δραστηριοτήτων. Επιπλέον, ο διαμορφωτής δέσμης LCMV υπερέχει ως προς την ικανότητά του να εντοπίζει τόσο σύγχρονες όσο και ασύγχρονες νευρικές δραστηριότητες, καθιστώντας τον ιδιαίτερα ευέλικτο [22]. Αυτή η ευελιξία είναι ζωτικής σημασίας για μελέτες που διερευνούν σύνθετες γνωστικές διαδικασίες ή διερευνούν τις αλληλεπιδράσεις μεταξύ διαφορετικών περιοχών του εγκεφάλου. Ο διαμορφωτής δέσμης LCMV είναι ιδιαίτερα κατάλληλος για μελέτες που απαιτούν τον εντοπισμό δυναμικών εγκεφαλικών διεργασιών. Η ικανότητά του να παρέχει μια χωρική χαρτογράφηση υψηλής ανάλυσης της εγκεφαλικής δραστηριότητας τον καθιστά ανεκτίμητο για την κατανόηση των ταχέων αλλαγών στα νευρωνικά δίκτυα κατά τη διάρκεια διαφόρων γνωστικών εργασιών. Για παράδειγμα, στην έρευνα της γνωστικής νευροεπιστήμης, ο διαμορφωτής δέσμης LCMV μπορεί να χρησιμοποιηθεί για τη μελέτη των εγκεφαλικών δραστηριοτήτων που εμπλέκονται στη λήψη αποφάσεων, την επεξεργασία της γλώσσας ή την εναλλαγή της προσογής.

Παρά τα πλεονεκτήματά της, η διαμόρφωση δέσμης αντιμετωπίζει ορισμένες προκλήσεις και περιορισμούς. Μια σημαντική πρόκληση είναι η ανάγκη για ακριβή μοντέλα κεφαλής. Η αποτελεσματικότητα της διαμόρφωσης δέσμης εξαρτάται σε μεγάλο βαθμό από την ακρίβεια αυτών των μοντέλων, τα οποία πρέπει να αναπαριστούν με ακρίβεια τις αγώγιμες ιδιότητες των διαφόρων ιστών του κεφαλιού. Οι ανακρίβειες σε αυτά τα μοντέλα μπορούν να οδηγήσουν σε σφάλματα στον εντοπισμό της πηγής. Μια άλλη πρόκληση στη διαμόρφωση δέσμης είναι η πιθανότητα διαρροής σήματος μεταξύ των περιοχών. Αυτό το φαινόμενο εμφανίζεται όταν η δραστηριότητα από μια περιοχή του εγκεφάλου αποδίδεται εσφαλμένα σε μια άλλη περιοχή, οδηγώντας σε παραπλανητικά αποτελέσματα. Η αντιμετώπιση της διαρροής σήματος απαιτεί προσεκτική βαθμονόμηση των χωρικών φίλτρων και μπορεί να περιλαμβάνει εξελιγμένες υπολογιστικές τεχνικές για τη διασφάλιση του ακριβούς εντοπισμού της πηγής.

## 2.2. Μηχανική και Βαθιά Μάθηση

Η Μηχανική Μάθηση (Machine Learning) αφορά αλγορίθμους που μαθαίνουν μοτίβα από δεδομένα προκειμένου να κάνουν προβλέψεις ή ταξινομήσεις, χωρίς ρητή προγραμματισμένη εντολή. Στα παραδοσιακά συστήματα μηχανικής μάθησης για ταξινόμηση σημάτων EEG εφαρμόζεται μια διαδικασία δύο βημάτων: αρχικά πραγματοποιείται μη αυτόματη εξαγωγή χαρακτηριστικών από τα σήματα (π.χ. στατιστικά γνωρίσματα, ενεργειακό φάσμα, κ.ά.) και στη συνέχεια αυτά τα χαρακτηριστικά δίνονται ως είσοδοι σε κάποιον ταξινομητή. Η απόδοση των μοντέλων εξαρτάται σε μεγάλο βαθμό από την ποιότητα των εξαγώμενων και χαρακτηριστικών (feature engineering) που προηγείται. Υπάρχει πληθώρα καθιερωμένων αλγορίθμων μηχανικής μάθησης για ταξινομήσεις EEG: ενδεικτικά αναφέρονται οι Υποστηρικτικές Μηγανές Διανυσμάτων (SVM), τα Δέντρα Απόφασης, τα Τυχαία Δάση (Random Forests),  $\eta$  Aoyiotiký Παλινδρόμηση, οι k-πλησιέστεροι Γείτονες (k-NN), o Naïve Bayes, καθώς και μέθοδοι ενσωμάτωσης όπως τα Boosted Trees (π.γ. AdaBoost) [23]. Οι κλασικοί αυτοί ταξινομητές συχνά επιτυγχάνουν ικανοποιητικά αποτελέσματα ακόμα και με περιορισμένα σύνολα δεδομένων, καθώς είναι λιγότερο επιρρεπείς σε υπερπροσαρμογή σε σχέση με μεγάλα νευρωνικά δίκτυα. Ένα σημαντικό πλεονέκτημά τους είναι η διαφάνεια και επεζηγησιμότητα: επειδή βασίζονται σε συγκεκριμένα, χειροποίητα γνωρίσματα και χαρακτηριστικά που έχουν προέλθει από το feature engineering, ο ειδικός μπορεί ευκολότερα να κατανοήσει ποιες μεταβλητές συντελούν στην απόφαση του μοντέλου. Αυτό το γαρακτηριστικό είναι ιδιαίτερα επιθυμητό σε ιατρικές εφαρμογές, όπου το να εξηγείται η απόφαση του συστήματος είναι συχνά εξίσου κρίσιμο με την ακρίβεια της πρόβλεψης.

Η **Βαθιά Μάθηση (Deep Learning)**, από την άλλη πλευρά, αποτελεί υποπερίπτωση της μηχανικής μάθησης που χρησιμοποιεί πολυεπίπεδα τεχνητά νευρωνικά δίκτυα για αυτόματη μάθηση σύνθετων αναπαραστάσεων των δεδομένων. Σε αντίθεση με τη συμβατική προσέγγιση όπου τα χαρακτηριστικά ορίζονται εκ των προτέρων, τα βαθιά νευρωνικά δίκτυα μπορούν να μάθουν απευθείας από τα ακατέργαστα δεδομένα (raw data) και να εξαγάγουν ιεραρχικά γνωρίσματα υψηλού επιπέδου χωρίς ανθρώπινη παρέμβαση. Ένα τυπικό βαθύ νευρωνικό δίκτυο έχει δομή εμπρόσθιας τροφοδότησης (feedforward) και αποτελείται από μια είσοδο, πολλές κρυφές στοιβάδες ενδιάμεσων νευρώνων, και μια έξοδο ταξινόμησης [24]. Τα δίκτυα feedforward βασίζονται στους νευρώνες ή αλλιώς Perceptrons.



Σχήμα 2.7 Πολυστρωματικό δίκτυο από perceptrons και συνάρτηση perceptron

Κάθε κρυφή στοιβάδα εφαρμόζει έναν μετασχηματισμό στα δεδομένα (μέσω βαρών και συναρτήσεων ενεργοποίησης), δημιουργώντας βαθμιαία πιο αφηρημένες αναπαραστάσεις. Μία ιδιαίτερα σημαντική κατηγορία είναι τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks, CNNs) [25], τα οποία αποτελούνται από συνελικτικά επίπεδα (convolutional layers), επίπεδα υποδειγματοληψίας/συγκέντρωσης (π.χ. max-pooling) και πλήρως συνδεδεμένα επίπεδα. Τα συνελικτικά φίλτρα ανιχνεύουν τοπικά χωροχρονικά μοτίβα στα δεδομένα. Οτρόπος είναι ότι τα φίλτρα συνέλιξης, τα οποία είναι εκπαιδεύσιμοι παράμετροι του συστήματος, παράγουν συνέλιξη με την αντίστοιχη είσοδο (βλ Σχήμα 2.8) – για παράδειγμα, σε εικόνες ή καταγραφές EEG τα επίπεδα συνελίκωσης μπορούν να μάθουν χαρακτηριστικές τοπικές δομές ή ρυθμούς [23].



Σχήμα 2.8 Διακρητή συνέλιξη δισδιάστατης εισόδου και φίλτρου

Τα επίπεδα pooling μειώνουν την ανάλυση/διάσταση διατηρώντας τις πιο σημαντικές πληροφορίες, παρέχοντας κάποια βαθμό αμεταβλητότητας σε μικρές μεταβολές του σήματος. Οι παραπάνω λειτουργίες δρουν ως αυτόματος εξαγωγέας χαρακτηριστικών, ενώ στο τελικό τμήμα του δικτύου συνήθως προστίθενται μία ή περισσότερες πλήρως διασυνδεδεμένες (fully-connected) στοιβάδες που λειτουργούν ως κλασικός ταξινομητής αξιοποιώντας τα εξαγμένα γνωρίσματα [24]. Τα βαθιά δίκτυα όπως τα CNN έχουν επιδείξει κορυφαίες επιδόσεις σε προβλήματα εικόνων, ήχου και σήματος, επειδή μαθαίνουν παράλληλα τόσο χωρικές όσο και χρονικές εξαρτήσεις στα δεδομένα. Ωστόσο, μειονέκτημά τους είναι ότι πρόκειται για «μαύρα κουτιά» – τα πρότυπα που μαθαίνουν είναι εσωτερικά περίπλοκα και δύσκολα ερμηνεύσιμα από άνθρωπο. Επιπλέον, απαιτούν μεγάλο όγκο δεδομένων για να γενικεύσουν σωστά (συχνά περισσότερα από ό,τι οι ρηχές μέθοδοι) και εκπαίδευση με υψηλό υπολογιστικό κόστος (π.χ. σε GPU) [23].

Κατά την εκπαίδευση ενός βαθιού νευρωνικού δικτύου, χρησιμοποιείται συνήθως αλγόριθμος διάδοσης του σφάλματος προς τα πίσω (backpropagation) σε συνδυασμό με μεθόδους βελτιστοποίησης (όπως παραλλαγές βαθμωτής καθόδου) για την προσαρμογή των βαρών σε κάθε επανάληψη. Για τη βελτίωση της διαδικασίας εκμάθησης και την αποφυγή υπερπροσαρμογής (overfitting), έχουν αναπτυχθεί διάφορες τεχνικές ομαλοποίησης (regularization) που ενσωματώνονται στην αρχιτεκτονική. Μία από τις πλέον διαδεδομένες είναι το **dropout**, όπου κατά την εκπαίδευση *απενεργοποιείται τυχαία ένα ποσοστό νευρώνων* σε κάθε στρώμα και κάθε επανάληψη [26]. Η ιδέα του dropout είναι ότι αποτρέπει τους νευρώνες από το να προσαρμοστούν υπερβολικά στα δεδομένα (*co-adapting*), αναγκάζοντας το δίκτυο να μάθει πιο γενικά πρότυπα – με αποτέλεσμα σημαντική μείωση του φαινομένου υπερπροσαρμογής και βελτίωση της ικανότητας γενίκευσης. Μια άλλη θεμελιώδης τεχνική είναι η **Κανονικοποίηση Κατά Συστοιχίες** (*Batch Normalization*), που εισήχθη από τους Ioffe και Szegedy (2015). Η μέθοδος αυτή κανονικοποιεί τις εξόδους κάθε κρυφής στοιβάδας ως προς τη μέση τιμή και διασπορά τους μέσα σε κάθε mini-batch εκπαίδευσης [27]. Με τον τρόπο αυτό αντιμετωπίζεται το πρόβλημα της «εσωτερικής μεταβολής κατανομών» (internal covariate shift) κατά την εκπαίδευση, επιτρέποντας την χρήση υψηλότερων ρυθμών μάθησης και καθιστώντας το δίκτυο λιγότερο ευαίσθητο στην αρχικοποίηση των βαρών. Επιπροσθέτως, η batch normalization δρα και ως μορφή κανονικοποίησης (regularizer), συχνά μειώνοντας την ανάγκη για άλλες τεχνικές όπως το dropout. Τέλος, στο επίπεδο εξόδου ενός νευρωνικού ταξινομητή χρησιμοποιείται συνήθως η συνάρτηση **softmax** σ(z) για πολυκλασικά προβλήματα. Η συνάρτηση softmax λαμβάνει ως εισόδους τους μηκανονικοποιημένους βαθμολογικούς δείκτες του μοντέλου (logits) και τους μετασχηματίζει σε πιθανότητες ανά κλάση, παράγοντας ένα διάνυσμα κανονικοποιημένων πιθανοτήτων που αθροίζει σε 1.

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$
(2.8)

Αυτό επιτρέπει την ερμηνεία της εξόδου του δικτύου ως κατανομή πιθανότητας για τις διαθέσιμες κατηγορίες.

#### 2.2.1 Ταξινόμηση και Μετρικές Αξιολόγησης

Στα προβλήματα ταξινόμησης, ειδικά στον χώρο της βιοϊατρικής όπου τα δεδομένα μπορεί να είναι δυσαναλογημένα (imbalanced) ή το διακύβευμα υψηλό, η αξιολόγηση της απόδοσης ενός μοντέλου απαιτεί περισσότερα από μια απλή μέτρηση. Η **Ακρίβεια (accuracy)**, οριζόμενη ως το ποσοστό των σωστών προβλέψεων επί του συνόλου των δειγμάτων, είναι μια βασική μετρική [28].

$$Accuracy = \frac{\sum True \ Positives + \sum True \ Negatives}{\sum Total \ number \ of \ cases}$$
(2.9)

Ωστόσο, η ακρίβεια μπορεί να αποδειχθεί παραπλανητική όταν οι κλάσεις δεν εμφανίζονται με ίσες αναλογίες – για παράδειγμα, ένα σύστημα που πάντα προβλέπει την κυρίαρχη κλάση θα εμφανίσει υψηλή ακρίβεια, αλλά δεν θα έχει πραγματική διακριτική ικανότητα [28]. Για τον λόγο αυτό, χρησιμοποιούνται συμπληρωματικές μετρικές που βασίζονται στο υπόδειγμα του πίνακα σύγχυσης (confusion matrix), λαμβάνοντας υπόψη ξεχωριστά τα αληθινά και ψευδή θετικά/αρνητικά αποτελέσματα. Η Ευαισθησία (sensitivity) ή Ανάκληση (recall) ορίζεται ως το ποσοστό των πραγματικά θετικών περιστατικών που ανιχνεύονται σωστά από το μοντέλο [28].

$$Sensitivity = \frac{\sum True \ Positives}{\sum Total \ number \ of \ Positives}$$
(2.10)

Η ευαισθησία ισοδυναμεί με το *True Positive Rate* (TPR) και σε ένα ιατρικό σενάριο αντιστοιχεί στην ικανότητα της μεθόδου να μην χάνει περιστατικά της υπό ανίχνευση κατάστασης (π.χ. να ανιχνεύει όλους τους επιληπτικούς παροξυσμούς). Η Ειδικότητα

(specificity), αντιστρόφως, είναι το ποσοστό των πραγματικά αρνητικών περιστατικών (π.χ. υγιείς περιπτώσεις) που σωστά ταξινομούνται ως αρνητικά [28] – ισοδυναμεί με το True Negative Rate.

$$Specificity = \frac{\sum True \ Negative}{\sum Total \ number \ of \ Negatives}$$
(2.11)

Μια υψηλή ειδικότητα σημαίνει ότι το σύστημα αποφεύγει τα ψευδώς θετικά συναγερμούς. Η Precision ή Θετική Προγνωστική Αξία (PPV) μετρά το ποσοστό των προβλεπόμενων ως θετικών δειγμάτων που είναι πράγματι θετικά [28]. Δηλαδή, από όλα τα περιστατικά που το μοντέλο σηματοδότησε ως θετικά, η precision δείχνει πόσα είναι όντως αληθινά – μια χαμηλή τιμή precision υποδηλώνει πολλά ψευδώς θετικά. Μια συνοπτική μετρική που συνδυάζει την ευαισθησία με την precision είναι ο F1-score, ο οποίος ορίζεται ως η αρμονική μέση τιμή των δύο [28]. Το F1 λαμβάνει τιμές από 0 έως 1, και τιμωρεί τις ακραίες ανισορροπίες ανάμεσα σε recall και precision – είναι 1 μόνο όταν και οι δύο είναι 1, ενώ μειώνεται αισθητά αν μία από τις δύο είναι χαμηλή. Γι' αυτόν τον λόγο το F1-score είναι χρήσιμο σε περιπτώσεις όπου θέλουμε μια ενιαία ένδειξη απόδοσης που λαμβάνει υπόψη τόσο τα ψευδώς αρνητικά όσο και τα ψευδώς θετικά σφάλματα.

Πέρα από τις παραπάνω μετρικές που ορίζονται σε ένα συγκεκριμένο κατώφλι απόφασης, σημαντικό ρόλο παίζουν και οι καμπύλες αξιολόγησης όπως η ROC (Receiver Operating Characteristic), οι οποίες απεικονίζουν την απόδοση του μοντέλου σε όλα τα πιθανά κατώφλια διάκρισης. Συνοπτικός δείκτης αυτής της καμπύλης είναι το AUROC (Area Under the ROC Curve) – δηλαδή το εμβαδόν κάτω από την καμπύλη ROC [29]. Το AUROC παρέχει ένα μέτρο της συνολικής διακριτικής ικανότητας του ταξινομητή ανεξαρτήτως κατωφλίου, ίσο με την πιθανότητα το μοντέλο να βαθμολογήσει υψηλότερα ένα τυχαίο θετικό δείγμα από ένα τυχαίο αρνητικό. Μια τιμή AUROC=0.5 αντιστοιχεί σε τυχαίες μαντεψιές (καμία διακριτική ικανότητα), ενώ AUROC=1.0 υποδηλώνει τέλεια κατάταξη όλων των θετικών πάνω από τα αρνητικά. Στην πράξη, υψηλές τιμές AUROC (π.χ. >0.8) υποδεικνύουν ότι ο ταξινομητής έχει ισχυρή ικανότητα διάκρισης μεταξύ των κλάσεων σε ευρύ φάσμα πιθανών κατωφλίων, κάτι ιδιαίτερα σημαντικό όταν η απόφαση μπορεί να ρυθμιστεί ώστε να επιτυγχάνεται επιθυμητή ισορροπία ευαισθησίας/ειδικότητας ανάλογα με το εκάστοτε κλινικό ζητούμενο.

## 2.3 Εργαλεία Επεξηγησιμότητας (XAI) – Έμφαση στο SHAP και το Gradient Explainer

Η εντυπωσιακή απόδοση των μεθόδων βαθιάς μάθησης συνοδεύεται από το μειονέκτημα της περιορισμένης διαφάνειας: τα μοντέλα συχνά λειτουργούν ως «μαύρα κουτιά», ιδίως σε περιπτώσεις όπως η ανάλυση βιοϊατρικών σημάτων όπου το δίκτυο εντοπίζει περίπλοκα χωροχρονικά πρότυπα που δεν είναι άμεσα διερμηνεύσιμα. Στις ιατρικές εφαρμογές, η επεξηγησιμότητα είναι κρίσιμη ώστε να κερδηθεί η εμπιστοσύνη των επαγγελματιών υγείας και να ενσωματωθούν οι προγνώσεις στη λήψη αποφάσεων [23]. Μάλιστα, όταν η δυνατότητα ερμηνείας θεωρείται υψίστης σημασίας, μπορεί να προτιμηθούν απλούστερα και πιο διαφανή μοντέλα μηχανικής μάθησης έναντι βαθιών νευρωνικών που επιτυγχάνουν ελαφρώς καλύτερη ακρίβεια [23]. Για να γεφυρωθεί αυτό το χάσμα μεταξύ **ακρίβειας** και **ερμηνευσιμότητας**, έχει αναδυθεί το πεδίο του **Εξηγήσιμου Τεχνητού Νοημού** 

(Explainable AI, XAI), το οποίο αναπτύσσει μεθόδους που επιτρέπουν στον άνθρωπο να κατανοήσει και να ερμηνεύσει τις αποφάσεις σύνθετων μοντέλων [30]. Τα εργαλεία ΧΑΙ διακρίνονται σε προσεγγίσεις a priori (σχεδιασμός εγγενώς διαφανών μοντέλων) και a posteriori (post hoc εξηγήσεις για ήδη εκπαιδευμένα μοντέλα). Στο πλαίσιο των post hoc μεθόδων – που είναι και οι πλέον κατάλληλες όταν δεν μπορούμε ή δεν θέλουμε να θυσιάσουμε την απόδοση των deep learning μοντέλων – έχουν προταθεί διάφορες τεχνικές. Για παράδειγμα, οι τοπικές προσεγγίσεις με μοντέλα-αντικαταστάτες όπως το LIME (Local Interpretable Model-Agnostic Explanations) εκπαιδεύουν ένα απλό, ερμηνεύσιμο μοντέλο (π.χ. ένα γραμμικό μοντέλο ή δέντρο) τοπικά γύρω από μια δεδομένη πρόβλεψη, προκειμένου να προσεγγίσουν τη συμπεριφορά του μαύρου κουτιού σε εκείνη τη γειτονιά του χώρου χαρακτηριστικών [31]. Συγκεκριμένα, το LIME παράγει πολλαπλές ελαφρώς τροποποιημένες εκδοχές μιας εισόδου (π.χ. τυχαίες διακυμάνσεις στα σήματα EEG) και παρατηρεί πώς αλλάζει η έξοδος του μοντέλου. Έπειτα εφαρμόζει ένα απλό μοντέλο πάνω σε αυτά τα δείγματα με βάρη που δίνουν έμφαση στα πιο κοντινά στην αρχική περίπτωση, ώστε το τοπικό υποκατάστατο να εξηγεί ποια γνωρίσματα της εισόδου συνέβαλαν περισσότερο στην απόφαση [31]. Άλλες προσεγγίσεις επεξηγησιμότητας βασίζονται στον υπολογισμό παραγώγων του μοντέλου ως προς την είσοδο – τα λεγόμενα gradient-based saliency maps. Για παράδειγμα, οι μέθοδοι Saliency/Gradient maps και Grad-CAM χρησιμοποιούν την ανάστροφη διάδοση (backpropagation) για να υπολογίσουν την συμβολή κάθε χαρακτηριστικού εισόδου στο τελικό σκορ, παράγοντας έναν «χάρτη» σημαντικότητας των εισόδων. Παρόμοια, η τεχνική Integrated Gradients [32] υπολογίζει την συμβολή ενός γαρακτηριστικού ενσωματώνοντας (ολικοποιώντας) την παράγωγο του δικτύου κατά μήκος μιας διαδρομής από μια ουδέτερη τιμή αναφοράς (baseline) μέχρι την πραγματική είσοδο, εξασφαλίζοντας ορισμένες ωφέλιμες θεωρητικές ιδιότητες (π.χ. συμμεταβλητότητα και ανεξαρτησία από την κλίμακα) για τις αποδόσεις σημαντικότητας που προκύπτουν.

Ανάμεσα στα εργαλεία ΧΑΙ, εξέχουσα θέση κατέχει το SHAP (SHapley Additive exPlanations) [30]. Το SHAP αποτελεί μια ενοποιημένη θεωρητική προσέγγιση για την επεξήγηση μεμονωμένων προβλέψεων, η οποία βασίζεται στην θεωρία παιγνίων συγκεκριμένα στις τιμές Shapley [33] που εισήγαγε ο Lloyd Shapley (1953) για τον καταμερισμό κέρδους σε συνεργατικά παιχνίδια. Στην πράξη, το SHAP αποδίδει σε κάθε χαρακτηριστικό εισόδου μια αξία σημαντικότητας (feature importance value) για μια συγκεκριμένη πρόβλεψη του μοντέλου. Οι Lundberg και Lee (2017) απέδειξαν ότι εάν απαιτήσουμε από τις μεμονωμένες συνεισφορές χαρακτηριστικών να πληρούν ορισμένες επιθυμητές ιδιότητες – όπως τοπική πιστότητα, μονοτονία/συνέπεια και γραμμικότητα – τότε υπάρχει μοναδική λύση για αυτές τις τιμές σημαντικότητας, η οποία αντιστοιχεί ακριβώς στις τιμές Shapley ενός κατάλληλου «παιχνιδιού» [30]. Με άλλα λόγια, το SHAP παρέγει τη θεωρητική εγγύηση ενός μοναδικού ορθολογικού τρόπου να μοιράσουμε την τελική πρόβλεψη ενός σύνθετου μοντέλου στα χαρακτηριστικά εισόδου του. Ένα πρακτικό αποτέλεσμα της θεωρίας είναι ότι το SHAP ενοποιεί πολλές προϋπάρχουσες μεθόδους ερμηνείας – οι Lundberg & Lee έδειξαν ότι αρκετές γνωστές τεγνικές (όπως LIME, DeepLIFT, Layer-Wise Relevance Propagation κ.ά.) αποτελούν ειδικές περιπτώσεις του πλαισίου SHAP ή προσεγγίζουν τις τιμές Shapley με διάφορους τρόπους. Λόγω αυτής της ενοποιητικής θεώρησης και της αποτελεσματικότητας των αλγορίθμων του, το SHAP έχει γνωρίσει ευρεία αποδοχή στην κοινότητα. Θεωρείται πλέον ένα από τα πιο αποτελεσματικά και αξιόπιστα εργαλεία XAI, κάτι που αποδεικνύεται από τον μεγάλο αριθμό αναφορών στη θεμελιώδη εργασία των δημιουργών του και τη γρήγορη ενσωμάτωσή του σε πληθώρα εφαρμογών. Ιδιαίτερα για δεδομένα χρονοσειρών και βιοσημάτων, το SHAP έχει δείξει ότι μπορεί να προσαρμοστεί κατάλληλα – μελέτες δείχνουν ότι μπορεί να αποδώσει σημαντικότητες χαρακτηριστικών σε τομείς όπως οι κατανομές ισχύος EEG στο χρόνο και τη συχνότητα, αναδεικνύοντας κρίσιμες περιοχές του σήματος που συμφωνούν και με γνωστούς βιοδείκτες [34]. Αυτό καθιστά το SHAP κατάλληλο για το πρόβλημα μας, όπου θέλουμε μετά την ταξινόμηση των σημάτων EEG να προσδιορίσουμε ποια χρονικά διαστήματα, συχνοτικές ζώνες ή εγκεφαλικές περιοχές συνέβαλαν περισσότερο στην απόφαση του μοντέλου.

Στην παρούσα διατριβή, όπου χρησιμοποιούνται βαθιά μοντέλα νευρωνικών δικτύων για την ταξινόμηση EEG, το βασικό εργαλείο επεξήγησης των αποτελεσμάτων είναι το πλαίσιο SHAP – συγκεκριμένα, η υλοποίηση Gradient Explainer <sup>1</sup> που παρέχεται από τη βιβλιοθήκη SHAP για βαθιά μάθηση. Ο Gradient Explainer είναι μια μέθοδος η οποία συνδυάζει την προσέγγιση των ολοκληρωμένων παραγώγων (integrated gradients) με τη θεωρία Shapley, ώστε να υπολογίζει αποτελεσματικά τις τιμές SHAP για πολύπλοκα διαφορίσιμα μοντέλα όπως τα νευρωνικά δίκτυα. Συγκεκριμένα, βασίζεται στην τεχνική των αναμενόμενων κλίσεων (expected gradients), όπου αντί να χρησιμοποιείται μια μοναδική ουδέτερη αναφορά όπως στα κλασικά integrated gradients, λαμβάνεται ένας μέσος όρος ως προς ένα σύνολο background τιμών εισόδου. Οι κλίσεις του μοντέλου υπολογίζονται πολλαπλές φορές μεταβαίνοντας βαθμιαία από αυτές τις τιμές-βάσεις προς την πραγματική είσοδο, και έπειτα γίνεται εξομάλυνση των αποτελεσμάτων. Έτσι, ο Gradient Explainer παράγει προσεγγιστικά τις τιμές Shapley για κάθε χαρακτηριστικό, οι οποίες αθροίζουν – σύμφωνα με την ιδιότητα προσθετικότητας και της γραμμικότητας – στη διαφορά μεταξύ της αναμενόμενης εξόδου του μοντέλου (για τυχαίες εισόδους-βάσεις) και της εξόδου για την τρέχουσα είσοδο. Με αυτόν τον τρόπο επιτυγχάνεται μια αποδοτική εκτίμηση των συνεισφορών των εισόδων στο αποτέλεσμα, ακόμη και για μεγάλα νευρωνικά δίκτυα, χωρίς να απαιτείται ο εξαντλητικός υπολογισμός όλων των πιθανών συνδυασμών χαρακτηριστικών (που είναι υπολογιστικά αδύνατος για πολλά χαρακτηριστικά). Ο Gradient Explainer αξιοποιείται στο πλαίσιο της εργασίας μας για να αποδώσει post hoc ερμηνείες στις ταξινομήσεις του συστήματος: μετά την απόφαση του δικτύου (π.χ. αν ένα σήμα EEG ανήκει σε μια συγκεκριμένη κατηγορία παθολογίας ή νοητικής κατάστασης), υπολογίζονται οι τιμές SHAP για όλα τα χαρακτηριστικά εισόδου. Τα χαρακτηριστικά αυτά μπορεί να είναι οι τιμές του σήματος σε συγκεκριμένες χρονικές στιγμές και συχνότητες (στην περίπτωση που χρησιμοποιείται αναπαράσταση γρόνου-συγνότητας) ή τα υπολογισμένα επίπεδα ενεργοποίησης ανά εγκεφαλική περιοχή (στην περίπτωση που εφαρμόζεται ανάλυση πηγών). Το αποτέλεσμα της μεθόδου είναι ότι σε κάθε μεμονωμένη πρόβλεψη μπορούμε να αποδώσουμε μια ποσοτική αιτιολόγηση: π.χ. συγκεκριμένα χρονικά τμήματα και συχνοτικές ζώνες εμφανίζονται με υψηλές θετικές τιμές SHAP, υποδεικνύοντας ότι η παρουσία αποφόρτισης σε αυτά αύξησε την πιθανότητα ταξινόμησης σε μια κλάση, ενώ άλλα στοιγεία μπορεί να έγουν αρνητικές τιμές SHAP, δείχνοντας ότι λειτούργησαν αποτρεπτικά για την εν λόγω κλάση. Επιπλέον, συγκεντρώνοντας τις τιμές SHAP σε πολλαπλά δείγματα (π.χ. μέσω μέσου όρου ή οπτικοποίησης κατανομών), μπορούμε να εξάγουμε και συμπεράσματα σε συνολικό επίπεδο σχετικά με το μοντέλο – για παράδειγμα, ποια χαρακτηριστικά του EEG (συγκεκριμένες περιοχές του εγκεφάλου ή συχνοτικά χαρακτηριστικά) είναι συνολικά τα πιο σημαντικά για τη διάκριση μεταξύ των τάξεων που μας ενδιαφέρουν. Αυτού του είδους οι επεξηγήσεις καθιστούν το σύστημα πιο αξιόπιστο και αποδεκτό, καθώς επιτρέπουν στον ερευνητή ή κλινικό ιατρό να επιβεβαιώσει ότι το μοντέλο βασίζεται σε φυσιολογικά εύλογα

 $<sup>^{1}\</sup> https://shap.readthedocs.io/en/latest/generated/shap.GradientExplainer.html$ 

μοτίβα (όπως γνωστούς βιοδείκτες) και όχι σε τυχαίο θόρυβο ή σφάλμα των δεδομένων. Το **SHAP με Gradient Explainer** επομένως ενισχύει την επεξηγησιμότητα των ταξινομήσεων EEG, παρέχοντας ένα παράθυρο ερμηνείας μέσα στο μαύρο κουτί του νευρωνικού δικτύου – κάτι ουσιώδες για την ασφαλή και αποτελεσματική ενσωμάτωση τέτοιων συστημάτων στην κλινική πράξη.

# 3. Κεφάλαιο 3 – Επεξηγησιμότητα με χρήση ανάλυσης χρόνου συχνότητας

Οι σχεδιασμοί Βαθιάς Μάθησης έχουν επανειλημμένα επιδείξει την αποτελεσματικότητά τους, ιδιαίτερα όσον αφορά την ικανότητά τους να ανιχνεύουν αυτόματα κρίσιμα χαρακτηριστικά από ακατέργαστα δεδομένα. Ωστόσο, αυτή η διαδικασία απαιτεί πολλαπλά επίπεδα βαθιών αρχιτεκτονικών, απαιτώντας κατά συνέπεια σημαντικούς υπολογιστικούς πόρους για την εκπαίδευση και τις προβλέψεις. Στην παρούσα εργασία, προτείνεται ένα επιφανειακό συνελικτικό νευρωνικό δίκτυο, με τρία συνελικτικά επίπεδα, για την ανίγνευση κόπωσης χρησιμοποιώντας δεδομένα ηλεκτροεγκεφαλογραφήματος (ΗΕΓ). Ως εκ τούτου, πραγματοποιήθηκε ανάλυση κυματιδίου Morlet στο πεδίο χρόνου-συχνότητας για την ανάλυση των δεδομένων και τον υπολογισμό των χαρακτηριστικών εισόδου για το μοντέλο Βαθιάς Μάθησης. Αυτά τα χειροποίητα χαρακτηριστικά συνδυάστηκαν στη συνέχεια με τα χαρακτηριστικά ανώτερου επιπέδου που έμαθε το μοντέλο, με αποτέλεσμα μια ανθεκτική λύση, μειώνοντας παράλληλα τον χρόνο εκπαίδευσης και το υπολογιστικό κόστος. Το προτεινόμενο μοντέλο κατάφερε να επιτύχει πολύ υψηλή ακρίβεια πρόβλεψης (97%) με ελάγιστες υπολογιστικές απαιτήσεις, ενώ η επακόλουθη ανάλυση τιμών SHAP παρείγε ενδείξεις για τα διακριτά χαρακτηριστικά των εγκεφαλικών κυμάτων που συνέβαλαν στη δημιουργία του μοντέλου. Από αυτή την άποψη, η παροχή προεπεξεργασμένων δεδομένων παρουσιάζει υψηλή συνάφεια όταν εφαρμόζεται σε μια αρχιτεκτονική βαθιάς μάθησης, ανοίγοντας τον δρόμο για συστήματα ανίχνευσης νοητικής κόπωσης που λειτουργούν σε πραγματικό χρόνο και προσδίδουν υψηλές δυνατότητες επεξηγησιμότητας.

## 3.1 Εισαγωγή

Η στέρηση ύπνου συνιστά μια κατάσταση που προκαλείται είτε από πλήρη έλλειψη ύπνου είτε από υποβέλτιστη διάρκεια ανάπαυσης. Μελέτες βασισμένες σε στοιχεία έχουν καταδείξει σημαντική επίδραση της στέρησης ύπνου στην επεξεργασία πληροφοριών και στη συνολική γνωστική απόδοση [35], με τις επιδράσεις της να ποικίλλουν ανάλογα με τις γνωστικές απαιτήσεις και τις εμπλεκόμενες εγκεφαλικές περιοχές. Υπό αυτή την κατάσταση, η νοητική κόπωση αναδύεται ως βασικό αποτέλεσμα, οδηγώντας σε δυσκολίες συγκέντρωσης και χαμηλή εγρήγορση. Οι συνέπειες είναι ιδιαίτερα εμφανείς στο πλαίσιο της εργαζόμενης μνήμης, η οποία εξ ορισμού συνεπάγεται ταχεία διαχείριση πληροφοριών και εκτέλεση εργασιών [36]. Ως εκ τούτου, η νοητική κόπωση σχετίζεται με μειωμένη απόδοση από την οπτική της εργαζόμενης μνήμης, η οποία συχνά αποδεικνύεται κρίσιμη σε εξειδικευμένα περιβάλλοντα όπως τα ιδρύματα υγειονομικής περίθαλψης. Συγκεκριμένα, το νοσοκομειακό προσωπικό συνήθως υφίσταται παρατεταμένες περιόδους απαιτητικής γνωστικής δραστηριότητας κατά τη διάρκεια των βαρδιών του, παρουσιάζοντας αυξημένη πιθανότητα επαγγελματικής εξουθένωσης, η οποία με τη σειρά της ενέχει πιθανό κίνδυνο για την ασφάλεια των ασθενών [37]. Υπό τέτοιες συνθήκες, είναι απαραίτητη η άμεση αναγνώριση των επιπέδων νοητικής κόπωσης που μπορεί να προκαλέσουν μειωμένη απόδοση και συναφή σφάλματα σε νοσοκομειακά περιβάλλοντα.

Για την αξιολόγηση των καταστάσεων νοητικής κόπωσης, έχουν χρησιμοποιηθεί διάφορες μέθοδοι ηλεκτροφυσιολογικής καταγραφής, παρέχοντας ενδείξεις για τις πολύπλοκες εγκεφαλικές αλληλεπιδράσεις και συνεπώς γνώσεις για τους γνωστικούς μηχανισμούς που

διέπουν τŋ γνωστική εξάντληση [38], [39]. Υπό αυτό το πρίσμα, το ηλεκτροεγκεφαλογράφημα (ΗΕΓ) έχει αναδειχθεί ως ένα ανεκτίμητο εργαλείο λόγω της ικανότητάς του να ανιχνεύει την εγκεφαλική δραστηριότητα (μη επεμβατικά) με τη μορφή νευρωνικών ταλαντώσεων, επιτρέποντας στους ερευνητές να ποσοτικοποιήσουν την επίδραση στις γνωστικές λειτουργίες υπό διάφορα επίπεδα νοητικού φόρτου [40]. Το κύριο πρόβλημα ταξινόμησης έγκειται στο εάν το άτομο βρίσκεται σε κατάσταση διανοητικής ανάπαυσης (εφεξής αναφερόμενη ως κλάση «Ανάπαυσης» - Rested class) ή διανοητικής κόπωσης (κλάση «Κόπωσης» - Fatigued class). Ως επί το πλείστον, τεχνικές επεξεργασίας σήματος επιστρατεύονται για την ανάπτυξη αλγορίθμων ανίχνευσης διανοητικής κόπωσης. Αυτές οι τεχνικές περιλαμβάνουν τη χρήση μαθηματικών αλγορίθμων για την ανάλυση και εξαγωγή σχετικών πληροφοριών από σήματα ΗΕΓ, χωρίς τη χρήση ανάπτυξης βασισμένης σε δεδομένα (data-driven). Για παράδειγμα, αρκετές μελέτες περιλαμβάνουν τον μετασχηματισμό κυματιδίων και την ανάλυση ανεξάρτητων συνιστωσών για την ανάλυση δεδομένων ΗΕΓ και τον εντοπισμό αλλαγών στην εγκεφαλική δραστηριότητα που είναι ενδεικτικές διανοητικής κόπωσης [41], [42]. Εντούτοις, αυτές οι προσεγγίσεις απαιτούν χειροκίνητη παρέμβαση και αναδιάρθρωση (refactoring), γεγονός που ενδέχεται να παρεμποδίζει την πλήρως αυτοματοποιημένη ανίχνευση της διανοητικής κόπωσης. Μια άλλη ευρέως χρησιμοποιούμενη προσέγγιση είναι η μηχανική μάθηση, η οποία περιλαμβάνει τη χρήση στατιστικών μοντέλων για την ανάλυση και ερμηνεία δεδομένων (ΗΕΓ) προκειμένου να εντοπιστούν πρότυπα και γαρακτηριστικά (π.γ., φάσμα ισγύος) που είναι ενδεικτικά της διανοητικής κόπωσης [43]. Εντούτοις, τέτοιες τεχνικές απαιτούν χειροκίνητη εξαγωγή χαρακτηριστικών για να αποδώσουν βέλτιστα, ενώ τα εξαγόμενα χαρακτηριστικά είναι συνήθως εξαρτώμενα από το υποκείμενο [43] και επομένως δεν είναι ιδανικά για παραδείγματα πραγματικού κόσμου και πραγματικού χρόνου.

Υπό αυτό το πρίσμα, η τεχνητή νοημοσύνη (TN) έχει καταστεί ένα ισχυρό εργαλείο, με τις αρχιτεκτονικές βαθιάς μάθησης (BM) (deep learning - DL) να επιτυγχάνουν αυτόματη ανίχνευση και εξαγωγή χαρακτηριστικών, επιφέροντας εξαιρετικές προόδους σε εργασίες ταξινόμησης, τμηματοποίησης και πρόβλεψης. Ως εκ τούτου, πολλές μελέτες ανίχνευσης κόπωσης έχουν χρησιμοποιήσει αλγορίθμους BM λόγω της υψηλής τους απόδοσης στην ταξινόμηση (σε σύγκριση με τις συμβατικές μεθόδους μηχανικής μάθησης), ειδικά όταν χρησιμοποιούν αποκλειστικά σήματα ΗΕΓ (παραβλέποντας άλλες τροπικότητες, όπως ηλεκτροοφθαλμογραφήματα ή δεδομένα βίντεο προσώπου) [44], [45]. Περαιτέρω, στην αναζήτηση μετρικών με υψηλή διακριτική ικανότητα, ακόμη πιο πολύπλοκες αρχιτεκτονικές έχουν προταθεί στη βιβλιογραφία, με συνελικτικά επίπεδα και αναδρομικά νευρωνικά δίκτυα να επιτυγχάνουν ελπιδοφόρα αποτελέσματα, ωστόσο με κόστος την αυξανόμενη πολυπλοκότητα [46]. Σε αυτό το πλαίσιο, οι περισσότερες από αυτές τις υλοποιήσεις χρησιμοποιούν πολυεπίπεδα βαθιά συνελικτικά νευρωνικά δίκτυα (CNNs) που απαιτούν υλισμικό (hardware) υψηλών προδιαγραφών και σημαντικό χρονικό διάστημα τόσο για τις φάσεις εκπαίδευσης όσο και πρόβλεψης. Παρόλο που τέτοιες διαμορφώσεις συνήθως επιτυγχάνουν υψηλή ακρίβεια ταξινόμησης, το συνολικό υπολογιστικό κόστος τις καθιστά ανεφάρμοστες για εφαρμογές πραγματικού κόσμου (π.χ., διεπαφές εγκεφάλου-υπολογιστή), καθώς ελάγιστοι υπολογιστικοί πόροι και χρόνος πρόβλεψης είναι απαραίτητα προκειμένου να επιτευχθεί λειτουργικότητα πραγματικού χρόνου.

Ένας άλλος σημαντικός παράγοντας που αποτελεί περιορισμό στις τρέχουσες αρχιτεκτονικές BM είναι η επεξηγησιμότητά τους. Η BM συνεπάγεται την ανάγκη για τεράστια, πολυεπίπεδα δίκτυα που είναι δύσκολο να εκπαιδευτούν και ακόμη δυσκολότερο να επεξηγηθούν. Τα μοντέλα γίνονται αντιληπτά ως «μαύρα κουτιά» με ελάχιστη κατανόηση της πραγματικής χρηστικότητας του συστήματος [47]. Η επεξηγησιμότητα είναι κρίσιμης σημασίας, καθώς αυξάνει την αξιοπιστία του συστήματος. Η κατανόηση του λόγου πίσω από τις προβλέψεις ενός μοντέλου μπορεί να προσδώσει εμπιστοσύνη στα αποτελέσματα του μοντέλου, παρέχοντας ταυτόχρονα ενδείξεις της ροής του συλλογισμού πίσω από τα παραγόμενα αποτελέσματα.

Λαμβάνοντας υπόψη τα παραπάνω, στην παρούσα μελέτη χρησιμοποιήθηκε μια αρχιτεκτονική BM, η οποία αξιοποίησε δεδομένα ΗΕΓ από μια δοκιμασία εργαζόμενης μνήμης που ενεργοποιούσε διαδικασίες οπτικής μνήμης. Οι συνολικές διαδικασίες περιελάμβαναν δύο καταγραφές (πριν και μετά από βάρδιες εφημερίας) ιατρικού και νοσηλευτικού προσωπικού, προκαλώντας υψηλή διανοητική κόπωση λόγω της επίδρασης της στέρησης ύπνου. Σε αυτό το πλαίσιο, εφαρμόστηκε μια σχετικά «ρηχή» (shallow) αρχιτεκτονική νευρωνικού δικτύου (με τρία επίπεδα διαδοχικών CNNs και ένα πλήρως συνδεδεμένο επίπεδο), βασισμένη σε μη επεμβατικές καταγραφές ΗΕΓ και κατάλληλη (προ-)ανάλυση δεδομένων. Τα αποτελέσματά μας υποδεικνύουν την αποτελεσματικότητα της προτεινόμενης μεθοδολογίας τόσο ως προς την απόδοση ταξινόμησης όσο και ως προς το υπολογιστικό κόστος. Αναλυτικότερα, η ροή εργασιών μας πέτυχε ακρίβεια 97% (97% Fscore) εντός ενός ιδιαίτερα οικονομικά αποδοτικού πλαισίου (cost-effective setup), φέροντας αξιοσημείωτες προοπτικές υλοποίησης για εφαρμογές πραγματικού χρόνου. Επιπλέον, η εφαρμογή ενός σχεδιασμού επεξηγησιμότητας μπόρεσε να παράσχει ενδείξεις των χαρακτηριστικών εγκεφαλικής δραστηριότητας εισόδου που χρησιμοποιήθηκαν από τον αλγόριθμο BM για την αξιολόγηση της διανοητικής κόπωσης.

## 3.2 Ανάλυση χρόνου-συχνότητας για την ανίχνευση ψυχικής κόπωσης

#### 3.2.1. Συμμετέχοντες

Στη μελέτη αυτή συμμετείχαν 22 υγιείς συμμετέχοντες (9 γυναίκες, μέση ηλικία 27,3 ± 4,1) που προσλήφθηκαν από το 401 Γενικό Στρατιωτικό Νοσοκομείο Αθηνών. Όλοι οι συμμετέχοντες ήταν γιατροί και μέλη του προσωπικού και ανέφεραν φυσιολογική ή διορθωμένη έως φυσιολογική όραση. Πριν από το πείραμα, όλοι οι συμμετέχοντες υποβλήθηκαν σε προκαταρκτικό έλεγχο για να διασφαλιστεί ότι δεν είχαν διαταραχή ύπνου, δεν είχαν ιστορικό ψυχικής νόσου ή ΔΕΠΥ και δεν είχαν λάβει μακροχρόνια φαρμακευτική αγωγή. Το πείραμα εγκρίθηκε από την Επιτροπή Ελέγχου του Ιδρύματος σύμφωνα με τη Διακήρυξη του Ελσίνκι, ενώ ελήφθη γραπτή συγκατάθεση μετά από ενημέρωση από όλα τα υποκείμενα.

#### 3.2.2. Πειραματικός σχεδιασμός

Οι συμμετέχοντες υποβλήθηκαν σε δοκιμασία μνήμης εργασίας δύο φορές, πριν από τις βάρδιες εφημερίας (έως 28 ώρες με λίγο ή καθόλου ύπνο-Αναπαυμένοι) και μετά τη λήξη της βάρδιας εργασίας (Κουρασμένοι). Ως εκ τούτου, για να εκτιμηθεί η επίδραση της

πνευματικής κόπωσης στην ικανότητα της μνήμης εργασίας, χρησιμοποιήθηκε ένα οπτικό έργο n-back υπό σχεδιασμό N = 2 [48]. Κατά τη διάρκεια αυτής της εργασίας, οι συμμετέχοντες κλήθηκαν να θυμηθούν και να συγκρίνουν την εικόνα που εμφανίστηκε σε 2 δοκιμές πριν από το τρέχον οπτικό ερέθισμα. Σε κάθε δοκιμή, μια εικόνα εμφανίζόταν σε μία από τις τέσσερις γωνίες της οθόνης, με τους συμμετέχοντες να υποδεικνύουν μεταξύ 1 από τις 4 πιθανές συνθήκες πατώντας το αντίστοιχο κουμπί. Κάθε συνθήκη απαιτούσε συγκρίσεις όσον αφορά το περιεχόμενο και τη θέση της εικόνας ως εξής: (α) ίδια εικόνα και ίδια θέση, (β) ίδια εικόνα (διαφορετική θέση), (γ) ίδια θέση (διαφορετική εικόνα) και (δ) καμία ομοιότητα (διαφορετική εικόνα, διαφορετική θέση). Το πείραμα περιελάμβανε 72 δοκιμές (4 συνθήκες εξισορροπημένες) και διήρκεσε περίπου 5 λεπτά. Τα οπτικά ερεθίσματα εμφανίζονταν για 3,5 δευτερόλεπτα, παρεμβάλλοντας έναν σταθερό σταυρό για 1 δευτερόλεπτο. Πριν από τις καταγραφές ΕΕG πραγματοποιήθηκαν δοκιμές εξάσκησης για να διασφαλιστεί η κατανόηση των ατόμων όσον αφορά την εκτέλεση της εργασίας.

#### 3.2.3. Λήψη δεδομένων και προεπεξεργασία

Οι καταγραφές δεδομένων ΗΕΓ πραγματοποιήθηκαν με τη χρήση καλύμματος ηλεκτροδίων 64 καναλιών (Biosemi, Activetwo System, Άμστερνταμ, Κάτω Χώρες), σύμφωνα με το πρότυπο σύστημα τοποθέτησης 10-20 με ρυθμό δειγματοληψίας 512 Hz. Τα διπολικά σήματα του ηλεκτροκουλόγραφου καταγράφηκαν από ηλεκτρόδια που τοποθετήθηκαν στον εξωτερικό κόγχο, καθώς και πάνω και κάτω από τους οφθαλμούς. Τα ακατέργαστα δεδομένα ΗΕΓ υποδειγματοληπτήθηκαν στα 256 Hz, φιλτράρονταν με ζωνοπερατό φίλτρο από 1 έως 40 Hz και αναφέρονταν εκ νέου στο μέσο όρο όλων των ηλεκτροδίων. Επιπρόσθετη διόρθωση των τεχνουργημάτων πραγματοποιήθηκε με τη χρήση της Ανάλυσης Ανεξάρτητων Συνιστωσών (ICA) και την απόρριψη των συνιστωσών που συσχετίζονταν σε μεγάλο βαθμό με τα σήματα του ηλεκτροοπτικογράφου [49]. Λόγω της σημαντικής μόλυνσης από τεχνουργήματα, τα δεδομένα από 2 συμμετέχοντες αποκλείστηκαν, με αποτέλεσμα να προκύψουν 20 άτομα για περαιτέρω ανάλυση. Στη συνέχεια, τα σήματα τμηματοποιήθηκαν σε δοκιμές και προσαρμόστηκαν σε σχέση με μια βασική γραμμή 100 ms πριν από το ερέθισμα. Χρησιμοποιήθηκε μια πρόσθετη αφαίρεση των δεδομένων των πρώτων 0,5 δευτερολέπτων της έναρξης του ερεθίσματος για να αμβλυνθούν οι επιρροές που οφείλονται στις επιδράσεις των ερεθισμάτων [50], με αποτέλεσμα 18 (ανά συνθήκη) εποχές 3 δευτερολέπτων με βάση τις δοκιμές. Το ηλεκτρόδιο Ιz αφαιρέθηκε από την ανάλυση για λόγους συμμετρίας, αφήνοντας 63 κανάλια για τη μετέπειτα ανάλυση. Προκειμένου να εξουδετερωθεί η συμβολή της γνωστικής επεξεργασίας άσχετης με τη μνήμη εργασίας, μόνο οι σωστές απαντήσεις συμπεριλήφθηκαν στην ακόλουθη ανάλυση, με αποτέλεσμα να προκύψουν 2401 δείγματα: 1234 δείγματα δεδομένων σε κόπωση και 1167 δείγματα δεδομένων σε ηρεμία. Η προεπεξεργασία υλοποιήθηκε στο Matlab 2022b (Mathworks Inc., Natick, MA, USA) χρησιμοποιώντας το εργαλείο EEGLAB [51].

#### 3.2.4. Εξαγωγή χαρακτηριστικών

Παρά τον σχετιζόμενο με το σκανδάλη χαρακτήρα των δεδομένων, αποφασίστηκε η ταυτοποίηση των καταστάσεων ψυχικής κόπωσης, ενώ παραλείφθηκαν οι πληροφορίες για τα οπτικά ερεθίσματα (έτσι αφαιρέθηκε το χρονικό πλαίσιο 0,5 δευτερολέπτων μετά το σκανδάλη, όπως αναφέρθηκε στην προηγούμενη ενότητα). Ωστόσο, μια τέτοια διαμόρφωση δημιουργεί ένα σύνολο δεδομένων που δεν είναι κλειδωμένο σε φάσεις. Αυτό θα μπορούσε να αποτελέσει σημαντικό εμπόδιο στην ανάλυση με χρονικό κλείδωμα, καθώς οι σημαντικές

χρονικές πληροφορίες του σήματος θα αγνοούνταν και θα χάνονταν. Αντίθετα, η ανάλυση χρονικής συχνότητας (TFR) μπορεί να αξιοποιήσει τη χρονική και φασματική ανάλυση για τις μη-φασικά κλειδωμένες δοκιμές [52]. Ως εκ τούτου, η ανάλυση TFR εφαρμόστηκε ως βασική ανάλυση πριν από την εκμάθηση τεχνητής νοημοσύνης για την αντιμετώπιση πιθανών ασυνεπειών δεδομένων που θα μπορούσαν να αποδειχθούν επιζήμιες για την απόδοση του μοντέλου πρόβλεψης.

Συγκεκριμένα, τα προεπεξεργασμένα δεδομένα αναλύθηκαν με τη χρήση σύνθετων κυματιδίων Morlet (Morlet wavelet) [53] δημιουργώντας το τελικό σύνολο δεδομένων χαρακτηριστικών. Η αξιοποίηση των κυματιδίων Morlet έγινε λόγω της εξέχουσας σημασίας τους στην ανάλυση χρόνου-συχνότητας, καθώς παρέχουν ίση διακύμανση στο χρόνο και στη συχνότητα [54]. Με αυτή την προσέγγιση, οι ακατέργαστες τιμές ισχύος για κάθε bin συχνότητας υπολογίστηκαν με ένα μεταβαλλόμενο χρονικό παράθυρο (που βασίζεται στη συχνότητα), χρησιμοποιώντας ένα Gaussian taper με καθορισμένο πλάτος. Το πλάτος του κωνικού Morlet επιλέχθηκε ως υπερπαράμετρος με τιμή 4, με αποτέλεσμα τη χρονική ανάλυση 13 χρονικών bins. Αυτή η ανάλυση παράγει έναν χάρτη φασματικής-χρονικής ενεργοποίησης ισχύος που χρησιμοποιείται ως είσοδος για το επόμενο μοντέλο DL. Ως εκ τούτου, δημιουργήθηκε ένας τρισδιάστατος πίνακας TFR (ανά υποκείμενο και δοκιμή) που ενσωματώνει τα χαρακτηριστικά ισχύος Morlet, με διαστάσεις [C×F×T], όπου C είναι ο αριθμός των καναλιών, F είναι τα bins συχνότητας και T είναι τα time bins. Ο αριθμός των καναλιών στην προτεινόμενη διαμόρφωση είναι 63, με 40 χωρίσματα συχνότητας (από 1-40 Hz, με παράθυρα συγνότητας 1 Hz) και 13 χωρίσματα χρόνου (από 0,5 s έως 3,5 s με χρονικό παράθυρο 0,25 s). Οι υπολογισμοί κυματιδίων Morlet πραγματοποιήθηκαν με τη χρήση του FieldTrip Toolbox [55].

#### 3.2.5. Μοντελοποίηση τεχνητής νοημοσύνης

#### 3.2.5.1. Μοντέλο βαθιάς μάθησης

Η αρχιτεκτονική τεχνητής νοημοσύνης αποτελούνταν από τρία δισδιάστατα επίπεδα συνελικτικής μάθησης σε διαδοχική σειρά (Σχήμα 3.1). Στο πρώτο στρώμα, διεξάγεται μια συνέλιξη 3 × 3 πυρήνων κατά κανάλι (63 κανάλια), με 32 φίλτρα εξόδου (κάθε φίλτρο παράγει έναν πίνακα εξόδου- έτσι, προκύπτουν 32 πίνακες του πρώτου στρώματος). Δεν πραγματοποιήθηκε καμία συμπλήρωση και χρησιμοποιήθηκε ένα τυπικό βήμα 1. Το μέγεθος του πυρήνα επιλέχθηκε ειδικά για να ενσωματώσει χαρακτηριστικά ζώνης συχνοτήτων 3 Hz και ζώνης χρόνου 0,75 sec. Το παράθυρο συχνοτήτων 3 Hz επιλέχθηκε καθώς είναι ένα ευρύ εύρος ζώνης που μπορεί να συσγετιστεί και επομένως να εξηγηθεί (καθώς είναι καλός διαιρέτης) με τις πιο συχνά χρησιμοποιούμενες ζώνες ΗΕΓ (δέλτα, θήτα, άλφα, βήτα και γάμμα). Όσον αφορά τη διάρκεια του χρονικού παραθύρου, επιλέχθηκε ένα χρονικό παράθυρο 750 ms με βάση παρόμοια πειραματικά παραδείγματα (για εργασίες μνήμης εργασίας 2 πίσω) που παρουσιάζουν μέσο χρόνο αντίδρασης περίπου 700 ms [56]. Το δεύτερο στρώμα αποτελείται από 16 φίλτρα εξόδου με μέγεθος πυρήνα 3 × 3, υπολογίζοντας επίσης τη συνέλιξη κατά κανάλι, αποτυπώνοντας χαρακτηριστικά υψηλότερου επιπέδου των 32 πινάκων εισόδου του πρώτου στρώματος. Το τελευταίο στρώμα συνέλιξης αποτελείται από 4 φίλτρα εξόδου με μέγεθος πυρήνα 4  $\times$  4. Συνήθως στις αρχιτεκτονικές DL, ένα στρώμα συγκέντρωσης χρησιμοποιείται μετά τα στρώματα συνέλιξης για τη μείωση της δειγματοληψίας του χώρου χαρακτηριστικών [57]. Ωστόσο, η μείωση της διάστασης των εξόδων δεν απαιτήθηκε στην προτεινόμενη προσέγγιση και, συνεπώς, δεν χρησιμοποιήθηκε στρώμα συγκέντρωσης. Δεδομένου ότι το αντικείμενο της παρούσας μελέτης περιλαμβάνει την εφαρμογή μεθόδων TN με ελάχιστους υπολογιστικούς πόρους, οι υπερπαράμετροι του δημιουργηθέντος ρηχού δικτύου επιλέχθηκαν για γρήγορο χρόνο εκτέλεσης ανά χρονικό βήμα κατά την εκπαίδευση. Στη συνέχεια, ο επιλεγμένος αριθμός στρωμάτων ήταν 3, λαμβάνοντας υπόψη το μεσαίο μέγεθος του συνόλου δεδομένων. Επιπλέον, τα μεγέθη των φίλτρων ακολουθούν ένα λογαριθμικό μοτίβο, μειώνοντας τον αριθμό των χαρακτηριστικών μετά από κάθε στρώμα στο μισό, μειώνοντας έτσι τον χρόνο εκπαίδευσης και εξαγωγής συμπερασμάτων. Ο βελτιστοποιητής, ο ρυθμός μάθησης, το μέγεθος δέσμης και η συνάρτηση ενεργοποίησης επιλέχθηκαν με βάση τις συνήθεις πρακτικές DL και τη δοκιμή με διάφορες τιμές. Ο βελτιστοποιητής που επιλέχθηκε ήταν ο Adam [58], ο αρχικός ρυθμός μάθησης επιλέχθηκε να είναι 0,001, το μέγεθος δέσμης ήταν 32 δείγματα και ως συνάρτηση ενεργοποίησης επιλέχθηκε η σιγμοειδής συνάρτηση.



Σχήμα 3.1 Σχηματική απεικόνιση του προτεινόμενου πλαισίου. Ο αστερίσκος (\*) συμβολίζει τη συνάρτηση συνέλιζης

Τέλος, χρησιμοποιήθηκε ένα πλήρως συνδεδεμένο στρώμα (FC) με 792 κόμβους εισόδου και 2 κόμβους εξόδου για κάθε μία από τις κλάσεις. Προκειμένου να εισαχθεί μη γραμμικότητα στο μοντέλο μας, χρησιμοποιήθηκε μια περιορισμένη λογιστική συνάρτηση (σιγμοειδής συνάρτηση) ως συνάρτηση ενεργοποίησης για τις πιθανότητες εξόδου. Μετά από κάθε στρώμα συνελικτικής ανάλυσης, χρησιμοποιήθηκε ένα στρώμα ομαλοποίησης δέσμης για την αύξηση της σταθερότητας και της ευρωστίας του δικτύου TN κατά τη διάρκεια των διαδικασιών εκπαίδευσης. Επιπλέον, για τη μείωση της υπερπροσαρμογής και τον χειρισμό της αβεβαιότητας χρησιμοποιήθηκε μια διαμόρφωση εγκατάλειψης με πιθανότητα μηδενικού στοιχείου 0,25. Το dropout είναι μια τεχνική κανονικοποίησης που μπορεί να μειώσει την αβεβαιότητα στα μοντέλα βαθιάς μάθησης με τυχαία απόρριψη νευρώνων κατά τη διάρκεια της εκπαίδευσης. Όλοι οι αλγόριθμοι υλοποιήθηκαν με προσαρμοσμένο κώδικα στη γλώσσα προγραμματισμού Python και στο πλαίσιο βαθιάς μάθησης PyTorch, έκδοση 1.13 [59].

#### 3.2.5.2. Εκπαίδευση βαθιάς μάθησης

Οι υπολογιστικοί πόροι που χρησιμοποιήθηκαν για την εκπαίδευση του προαναφερθέντος μοντέλου αποτελούνται από έναν επεξεργαστή Intel (R) Core (TM) i7-9700 με 3,00 GHz και 16 GB μνήμης RAM 2667 MHz σε λειτουργικό σύστημα Windows 11 Pro έκδοση 22H2 (Microsoft, Redmond, Washington, ΗΠΑ). Δεν χρησιμοποιήθηκε GPU συμβατή με CUDA, οπότε η εκπαίδευση ΑΙ διεξήχθη εξ ολοκλήρου στην CPU. Το σύνολο δεδομένων γαρακτηριστικών που δημιουργήθηκε αποτελείται από 2401 δοκιμές: 1234 δείγματα για την κατηγορία Fatigued και 1167 για την κατηγορία Rested. Παρόλο που όλα τα υποκείμενα ολοκλήρωσαν τον ίδιο αριθμό εργασιών για κάθε κατάσταση, για τη δημιουργία του συνόλου δεδομένων λήφθηκαν υπόψη μόνο οι σωστές απαντήσεις, εισάγοντας μια μικρή ανισορροπία τάξεων. Ωστόσο, οι διαφορές ήταν αμελητέες και, στο εξής, δεν μπορεί να υποτεθεί ότι συνέβαλαν στην αντίστοιχη μεροληψία. Για την εκπαίδευση του σχεδιασμένου μοντέλου μάθησης, βαθιάς χρησιμοποιήθηκε μια στρατηγική εκπαίδευσης συνδυασμένων υποκειμένων, προκειμένου να αξιοποιηθούν χαρακτηριστικά που παρέχουν γενικευμένες (ανεξάρτητες από το υποκείμενο) πληροφορίες σχετικά με την ψυχική κόπωση. Ως εκ τούτου, όλες οι παρατηρήσεις των σημείων δεδομένων κόπωσης και ανάπαυσης όλων των υποκειμένων συνδυάστηκαν και χωρίστηκαν τυχαία στα τρία υποσύνολα δεδομένων, δηλαδή στα υποσύνολα εκπαίδευσης, επικύρωσης και δοκιμής. Αναλυτικότερα, το πλήρες σύνολο δεδομένων χωρίστηκε σε ένα υποσύνολο εκπαίδευσης (που περιείχε το 75% του πλήρους συνόλου δεδομένων- 925 δείγματα της κατηγορίας Fatigued και 875 της κατηγορίας Rested), ένα σύνολο επικύρωσης (που περιείχε το 15% του πλήρους συνόλου δεδομένων- 185 της κατηγορίας Fatigued και 175 της κατηγορίας Rested) και ένα σύνολο δοκιμής (που περιείχε το 10% του πλήρους συνόλου δεδομένων- 124 της κατηγορίας Fatigued και 117 της κατηγορίας Rested). Η ανάθεση έγινε τυχαία για να αποφευχθεί οποιαδήποτε μεροληψία επιλογής μεταξύ των συμμετεχόντων. Το μοντέλο νευρωνικού δικτύου εκπαιδεύτηκε επανειλημμένα χρησιμοποιώντας το σύνολο δεδομένων εκπαίδευσης για συγκεκριμένο αριθμό εποχών. Κάθε εποχή (η οποία είναι μια υπερπαράμετρος της διαδικασίας εκπαίδευσης) είναι ένα πλήρες πέρασμα ολόκληρου του συνόλου δεδομένων εκπαίδευσης από τον αλγόριθμο μάθησης. Αυτό απαιτείται για ένα μοντέλο μέχρι να συγκλίνει η διαδικασία μάθησης. Συνήθως, ο αριθμός των εποχών είναι εκατοντάδες ή χιλιάδες. Ωστόσο, το μοντέλο μας συγκλίνει σε σχεδόν 150 εποχές.

## 3.3 Αποτελέσματα μοντέλου χρόνου-συχνότητας (ανίχνευση ψυχικής κόπωσης)

#### 3.3.1 Επιδόσεις

Τα αποτελέσματα της συνολικής απόδοσης του μοντέλου ΤΝ παρουσιάζονται παρακάτω (Πίνακας 1, Σχήμα 3.2), όπως μετρήθηκαν με τέσσερις τυπικές μετρικές μηχανικής μάθησης: ακρίβεια, ακρίβεια, ανάκληση και F-score. Ως εκ τούτου, η επιτευχθείσα ακρίβεια ταξινόμησης ήταν 97%, με ένα επιπλέον πολύ υψηλό F-score (που έφτασε το 97%), υποδεικνύοντας έτσι ότι η ψυχική κόπωση αναγνωρίστηκε αποτελεσματικά (ο πίνακας σύγχυσης των αποτελεσμάτων του μοντέλου ΤΝ παρουσιάζεται στο Σχήμα 3.3α). Για τις 150 εποχές του μοντέλου μας, η εκπαίδευση ολοκληρώθηκε σε 44 λεπτά. Στο Σχήμα 3β παρουσιάζεται το ιστορικό επικύρωσης με τη μορφή της ακρίβειας επικύρωσης κατά τη διάρκεια των εποχών. Επιπλέον, ο εκτιμώμενος χρόνος επεξεργασίας για τον προσεγγιστικό χρόνο πρόβλεψης μιας μεμονωμένης περίπτωσης ήταν περίπου 4 ms (ακολουθώντας την

προτεινόμενη μεθοδολογία και τη ρύθμιση των υπολογιστικών πόρων που περιγράφεται παραπάνω).

Για να εκτιμηθεί κατά πόσον η προκύπτουσα απόδοση θα μπορούσε να βελτιωθεί περαιτέρω, αξιολογήθηκε ένα επιπλέον μοντέλο 300 εποχών. Ωστόσο, παρά το γεγονός ότι σε μια ξεχωριστή συνεδρία εκπαίδευσης η συνολική ακρίβεια αυξήθηκε σε 97,6 (»0,1% αύξηση), το σύνολο δεδομένων δοκιμής απαιτούσε διπλάσιο χρόνο εκπαίδευσης στην ίδια υπολογιστική διάταξη. Ως εκ τούτου, ο επιπλέον χρόνος θα μπορούσε να θεωρηθεί περιττός, δεδομένου ότι η ακρίβεια επικύρωσης φτάνει σε ένα οροπέδιο και το μοντέλο συγκλίνει (Σχήμα 3.3β).



Σχήμα 3.2 Οι επιδόσεις του προτεινόμενου μοντέλου όσον αφορά: (a) απώλεια εκπαίδευσης για τις εποχές- (β) απώλεια επικύρωσης για τις εποχές- (γ) χαρακτηριστική καμπύλη λειτουργίας δέκτη (ROC).



Σχήμα 3.3 (α) πίνακας σύγχυσης του προτεινόμενου μοντέλου- (β) καμπύλη επικύρωσης για 150 εποχές (μπλε) και 300 εποχές (ματζέντα). Τα πιο σκούρα χρώματα αντιπροσωπεύουν μια πιο ομαλή καμπύλη.

Πίνακας 1 Οι επιδόσεις του προτεινόμενου μοντέλου

|                         | Precision | Recall | F-score | Support (No. of Samples) |
|-------------------------|-----------|--------|---------|--------------------------|
| Fatigued                | 98%       | 97%    | 97%     | 124                      |
| Rested                  | 97%       | 97%    | 97%     | 117                      |
| <b>Overall Accuracy</b> | -         | -      | 97%     | 241                      |

Στις εφαρμογές τεχνητής νοημοσύνης, το υπολογιστικό κόστος είναι ένας σημαντικός παράγοντας για την εξισορρόπηση της απόδοσης και του χρόνου που απαιτείται για την επίλυση του μοντέλου. Από την άποψη αυτή, οι απαιτήσεις του προτεινόμενου μοντέλου μπορούν να θεωρηθούν ελάγιστες, τόσο όσον αφορά την υπολογιστική ισχύ όσο και τον χρόνο που απαιτείται για την εκπαίδευση και την αξιολόγηση του μοντέλου ΤΝ. Συγκεκριμένα, όπως περιγράφεται παραπάνω, το μοντέλο εκπαιδεύτηκε σε 44 λεπτά χωρίς πυρήνες CUDA, ενώ κάθε επακόλουθη πρόβλεψη πραγματοποιήθηκε σε 4 ms. Με βάση αυτή την παραδοχή, για να δοθεί μια ένδειξη της υπολογιστικής και της χρονικής αποδοτικότητας της προτεινόμενης διαμόρφωσης, πραγματοποιήθηκε μια ενδελεχής σύγκριση στις ίδιες ρυθμίσεις υπολογισμού. Ως εκ τούτου, αναδημιουργήθηκαν και εκπαιδεύτηκαν με την ίδια υπολογιστική ρύθμιση διάφορα σύγχρονα μοντέλα ΤΝ που εξετάζουν την ταξινόμηση της ψυχικής κόπωσης, με την αξιολόγηση της αλγοριθμικής αποδοτικότητας να περιλαμβάνει τον απαιτούμενο χρόνο για την εκπαίδευση και τη δοκιμή (Πίνακας 2). Συγκεκριμένα, τα αναδημιουργημένα μοντέλα ΤΝ τροφοδοτήθηκαν με τα ίδια δεδομένα εισόδου με την προτεινόμενη προσέγγιση και μετρήθηκε η διάρκεια της εκπαίδευσης και ο χρόνος πρόβλεψης. Τα μοντέλα περιελάμβαναν: (α) ένα διπλό νευρωνικό δίκτυο συνελικτικής ανάλυσης (CNN) [45], β) ένα μονοδιάστατο u-δίκτυο σε συνδυασμό με ένα δίκτυο μακράς βραχυπρόθεσμης μνήμης (LSTM) [46] και γ) ένα τροποποιημένο δίκτυο ανάλυσης κύριων συνιστωσών (PCANet) με έναν ταξινομητή μηχανής διανυσμάτων υποστήριξης (SVM) [60].

| AI Model                 | Authors | Dataset (Training Strategy)                                       | Classification<br>Accuracy |
|--------------------------|---------|---|----------------------------|
| UNET(CNN) +<br>LSTM      | [46]    | 9 subjects with 1 channel, 4 s segments (Subject-<br>specific)    | 83%                        |
| SVM                      | [61]    | 8 subjects with 19 channels, 10 s segments (Combined-<br>subject) | 89%                        |
| Dual CNN                 | [45]    | 22 subjects with 1 channel, 1 s segments (Combined-<br>subject)   | 93%                        |
| PCANet + SVM             | [60]    | 6 subjects with 32 channels, 4 s segments (Subject-<br>specific)  | 96%                        |
| This study (TF<br>+ CNN) | -       | 22 subjects with 64 channels, 3 s segments (Combined-<br>subject) | 97%                        |

Πίνακας 2 Σύγκριση κόστους-αποτελεσματικότητας αλγορίθμων.

Η είσοδος κάθε ενός από αυτά τα μοντέλα ήταν μια τροποποίηση του συνόλου δεδομένων που περιγράφεται στην ενότητα 2.4, προκειμένου να συμμορφωθεί με τις παραμέτρους και τους σχεδιασμούς που παρείχαν οι συγγραφείς των μελετών. Αναλυτικότερα, τα μοντέλα dual CNN και UNET(CNN) + LSTM απαιτούν χρονοσειρές ενός καναλιού και, ως εκ τούτου, δόθηκε μια χρονοσειρά ενός καναλιού. Ομοίως, η είσοδος για το PCANet ήταν ένας πίνακας  $C \times r$ , όπου C είναι ο αριθμός των καναλιών και r είναι ο αριθμός των συνιστωσών PCA των χρονοσειρών. προαναφερθέντα δεδομένων των μειωμένων Τα σγέδια TN αναδημιουργήθηκαν με τη χρήση του PyTorch για την ελαχιστοποίηση τυχόν διαφορών στην υλοποίηση. Είναι ενδιαφέρον ότι, παρόλο που το διπλό CNN και το UNET(CNN) + LSTM απαιτούν μόνο ένα κανάλι ως είσοδο, ο χρόνος που απαιτείται για την εκπαίδευση είναι πολλαπλάσιος από αυτόν που απαιτείται για το προτεινόμενο μοντέλο TN.

Σημειωτέον ότι οι πρόσθετες πληροφορίες σχετικά με την αλγοριθμική απόδοση εξαιρέθηκαν από την αξιολόγηση της σχέσης κόστους-αποτελεσματικότητας, δεδομένου ότι το επίκεντρο της παρούσας εργασίας είναι η αξιολόγηση της απόδοσης αρκετά ρηχών αρχιτεκτονικών νευρωνικών δικτύων με βάση τον κατάλληλο χειρισμό δεδομένων και δεν πρόκειται για σύγκριση με υπάρχοντα αλγοριθμικά σχέδια.

#### 3.3.2. Επεξηγησιμότητα

Ένα από τα σημαντικότερα ζητήματα που αφορούν τους σχεδιασμούς DL AI είναι η έμφυτη φύση τους ως «μαύρου κουτιού», γεγονός που καθιστά δύσκολη την εκτίμηση των σγέσεων μεταξύ των χαρακτηριστικών εισόδου, η οποία παρέχει αιτιολογία πίσω από τα αποτελέσματα που υπολογίζονται. Από αυτή την άποψη, ο κολοσσιαίος αριθμός στρωμάτων που περιέχονται στην πλειονότητα των αρχιτεκτονικών DL εμποδίζει τη σημασία κάθε γαρακτηριστικού που υπολογίζεται εσωτερικά (καθώς κάθε στρώμα μετασγηματίζει τα δεδομένα εισόδου). Ως αποτέλεσμα, η ερμηνευσιμότητα και, επομένως, η αξιοπιστία αυτών των μοντέλων μπορεί να διακυβευθεί. Τα ρηχά τεχνητά δίκτυα αμβλύνουν αυτές τις ανησυχίες, δεδομένου ότι ο σχετικά μικρός αριθμός στρωμάτων επιτρέπει την εκτίμηση και συνεπώς την εξήγηση των αποφάσεων που ενσωματώνονται σε κάθε βήμα του μοντέλου. Από την άποψη αυτή, τα εργαλεία επεξηγηματικότητας, όπως οι προσθετικές εξηγήσεις Shapley (SHAP), οι αξίες παρέχουν μια οπτικοποίηση των εσωτερικών διαδικασιών του μοντέλου [30], τονίζοντας τη σημασία κάθε περιοχής ή χαρακτηριστικού. Στην παρούσα μελέτη, η εκτίμηση των τιμών SHAP πραγματοποιήθηκε με τη χρήση μιας βελτιωμένης μεθόδου DeepLift (βαθιά εκμάθηση σημαντικών χαρακτηριστικών) [62] με το εργαλείο DeepExplainer Tool. Το DeepExplainer χρησιμοποιεί μια τεχνική δειγματοληψίας για την προσέγγιση των τιμών Shapley για ένα μοντέλο DL. Συγκεκριμένα, δημιουργείται ένας μεγάλος αριθμός τυχαίων δειγμάτων από το χώρο εισόδου του μοντέλου και με τη χρήση των δειγμάτων που προκύπτουν για την εκτίμηση των τιμών Shapley. Στην παρούσα μελέτη, χρησιμοποιήθηκαν 100 δείγματα για τη δημιουργία των τιμών SHAPley. Για κάθε μεμονωμένη είσοδο/παρατήρηση, οι τιμές SHAP υπολογίζονται για κάθε μία από τις κλάσεις εξόδου, με τις θετικές τιμές να υποδηλώνουν τη συμβολή στην επιλογή μοντέλου της αντίστοιχης κλάσης, ενώ οι αρνητικές τιμές υποδηλώνουν τον ανασταλτικό χαρακτήρα των χαρακτηριστικών. Στην παρούσα μελέτη, οι τιμές SHAP χρησιμοποιήθηκαν για να προσδιορίσουν σε τι στηρίζεται το μοντέλο (αυξάνοντας την πιθανότητα ο χάρτης χαρακτηριστικών εισόδου να είναι η προβλεπόμενη κλάση). Ένα παράδειγμα σχετικά με την παρατήρηση των θετικών και αρνητικών τιμών SHAP παρουσιάζεται στο Σχήμα 4. Είναι ενδιαφέρον ότι στην πλειονότητα των υπολογισμών των τιμών SHAP, το εύρος μεταξύ 5-15 Hz συμβάλλει θετικά στην πρόβλεψη της κλάσης Fatigued.



Σχήμα 3.4 Ένα παράδειγμα αναπαράστασης των τιμών SHAP. Οι είσοδοι υποδηλώνουν τους πίνακες TFR, ενώ η αντιστοιχία του μοντέλου αντικατοπτρίζει τις επιδράσεις των χρωματισμένων περιοχών σε κάθε συγκεκριμένη έζοδο κλάσης. Η επάνω σειρά αντιστοιχεί στην κλάση Fatigue ενώ η κάτω σειρά στην κλάση Rested

Κατά παρόμοιο τρόπο, η ενεργοποίηση στο εύρος 3-7 Hz στο χάρτη χρόνου-συχνότητας υποδηλώνει έναν ανασταλτικό παράγοντα για την κλάση Fatigued και θετικές τιμές για την κλάση Rested. Η σημασία της οπτικής ερμηνείας των τιμών SHAP έρχεται μαζί με την οπτική επιθεώρηση των χαρτών ισχύος TFR. Η οπτική ερμηνεία των χαρτών ισχύος TFR είναι ένα ανεκτίμητο εργαλείο για τους νευροεπιστήμονες, διευκολύνοντας τη συλλογιστική πίσω από τις προβλέψεις του μοντέλου AI και ρίχνοντας φως στην κατανόηση της εγκεφαλικής δραστηριότητας, των καταστάσεων του ΗΕΓ, των καταστάσεων και των ψυχικών ασθενειών [63].

## 3.4 Συζήτηση

Στην παρούσα μελέτη, αξιοποιήθηκε και δοκιμάστηκε ένα αρκετά ρηχό νευρωνικό δίκτυο για τη διάκριση μεταξύ της κατάστασης κόπωσης και της κατάστασης ανάπαυσης, σε ένα πείραμα μνήμης εργασίας EEG. Η προτεινόμενη αρχιτεκτονική ήταν σε θέση να επιτύχει υψηλή ακρίβεια ταξινόμησης (97%), μέσα σε ένα υπολογιστικά ανέξοδο πλαίσιο, τονίζοντας τη χρησιμότητα της ανάλυσης δεδομένων που απαιτείται για εφαρμογές πραγματικού χρόνου/πραγματικού κόσμου.

Σε πολλές περιπτώσεις, όπως η όραση υπολογιστών, τα κρυφά χαρακτηριστικά που εξήχθησαν από αρχιτεκτονικές DL ήταν πιο αποτελεσματικά σε ορισμένες εργασίες από τα χαρακτηριστικά που εξήχθησαν χειροκίνητα, γεγονός που εξηγεί γιατί οι προσεγγίσεις DL υπερτερούν έναντι αυτών με μηχανική μάθηση. Ωστόσο, αυτό δεν ισχύει σε όλες τις εργασίες. Αν και, σε ορισμένες περιπτώσεις, η αξιοποίηση ακατέργαστων δεδομένων ΕΕG με CNN επιτυγχάνει πολύ υψηλή ακρίβεια [45], η τροφοδοσία του δικτύου με αναλυμένα δεδομένα για την ταξινόμηση σήματος ΕΕG αποδεικνύεται η βέλτιστη οδός. Από την άλλη πλευρά, τα χειροποίητα χαρακτηριστικά (όπως αυτά που προτείνονται στην παρούσα μελέτη) χρησιμοποιούνται συνήθως με κλασικές τεχνικές μηχανικής μάθησης, καθώς δεν απαιτείται η αυτόματη εξαγωγή χαρακτηριστικών (το κύριο πλεονέκτημα των αρχιτεκτονικών βαθιάς

μάθησης). Στην περίπτωση αυτή, ο υπολογισμός των χαρακτηριστικών υψηλότερου επιπέδου παρουσιάζει ένα εγγενές πλεονέκτημα, καθώς το προτεινόμενο μοντέλο υπερτερεί έναντι των κλασικών τεχνικών ML όσον αφορά το F-score [64].

Ένας άλλος σημαντικός παράγοντας στις προσεγγίσεις μηχανικής μάθησης είναι οι στρατηγικές εκπαίδευσης που εφαρμόζονται για τη δημιουργία του μοντέλου ΤΝ. Συνήθως, η πλειοψηφία των παρόμοιων μελετών έχει χρησιμοποιήσει μια θεματική προσέγγιση [46], [60]. Η ειδική για κάθε υποκείμενο στρατηγική περιλαμβάνει τα σύνολα εκπαίδευσης και δοκιμής που πρέπει να χωριστούν για κάθε υποκείμενο ξεχωριστά, δημιουργώντας Ν αποτελέσματα (όπου Ν είναι ο αριθμός των υποκειμένων). Η τελική ακρίβεια του μοντέλου υπολογίζεται από τον μέσο όρο των αποτελεσμάτων όλων των υποκειμένων. Ωστόσο, παρόλο που αυτή η στρατηγική ειδικά για κάθε υποκείμενο αποδίδει συνήθως υψηλές επιδόσεις (δεδομένου ότι τα μοντέλα ΤΝ είναι ικανά να καταγράφουν εξατομικευμένα γαρακτηριστικά), η προσέγγιση αυτή δεν είναι ιδανική για τον εντοπισμό γενικευμένων γαρακτηριστικών που είναι εφαρμόσιμα, παρά τις παραλλαγές των υποκειμένων. Στην παρούσα μελέτη, εφαρμόστηκε μια υψηλή καθολική (ανεξάρτητη από το υποκείμενο) απόδοση και μια στρατηγική εκπαίδευσης συνδυασμένων υποκειμένων. Ως εκ τούτου, με τη συμπερίληψη δειγμάτων από όλα τα υποκείμενα στο σύνολο δεδομένων εκπαίδευσης, ελαχιστοποιείται η πιθανή δειγματοληπτική μεροληψία στα δεδομένα, καθιστώντας την εκπαίδευση του μοντέλου ΤΝ πιο ισχυρή και λιγότερο επιρρεπή σε σφάλματα [65].

Για να καταδειχθεί η εξέχουσα θέση του σχεδιασμού του χειροποίητου χαρακτηριστικού DL, μια συνολική σύγκριση μεταξύ των σύγχρονων μεθόδων TN που χρησιμοποιούνται στη βιβλιογραφία και της προτεινόμενης μεθοδολογίας παρουσιάζεται στον Πίνακα 3. Σημειωτέον ότι η σύγκριση περιλαμβάνει μελέτες στις οποίες εφαρμόζονται διαφορετικά σύνολα δεδομένων και αναλύσεις σε κάθε μέθοδο. Ως εκ τούτου, κάθε ένα από τα μοντέλα που παρουσιάζονται αντιμετωπίζει διαφορετικά προβλήματα, τομείς και χώρους χαρακτηριστικών. Θα πρέπει να τονιστεί ότι η απόδοση κάθε μοντέλου δεν βασίζεται μόνο στη μεθοδολογική αρχιτεκτονική αλλά και στη διαθεσιμότητα, την ποιότητα και τη μεταβλητότητα των δεδομένων. Ωστόσο, τα συνολικά αποτελέσματα υποδηλώνουν ότι το προτεινόμενο πλαίσιο παρέχει υψηλά αποτελέσματα στη διάκριση της ψυχικής κόπωσης (υπό μια γενικευμένη διαδικασία εκπαίδευσης), συγκρίσιμα με αυτά που παρουσιάζονται στη βιβλιογραφία.

| AI Model                 | Authors | Dataset (Training Strategy)                                       | Classification<br>Accuracy |
|--------------------------|---------|---|----------------------------|
| UNET(CNN) +<br>LSTM      | [46]    | 9 subjects with 1 channel, 4 s segments (Subject-<br>specific)    | 83%                        |
| SVM                      | [61]    | 8 subjects with 19 channels, 10 s segments (Combined-<br>subject) | 89%                        |
| Dual CNN                 | [45]    | 22 subjects with 1 channel, 1 s segments (Combined-<br>subject)   | 93%                        |
| PCANet + SVM             | [60]    | 6 subjects with 32 channels, 4 s segments (Subject-<br>specific)  | 96%                        |
| This study (TF<br>+ CNN) | -       | 22 subjects with 64 channels, 3 s segments (Combined-<br>subject) | 97%                        |

Πίνακας 3 Σύγκριση μεθόδων ταξινόμησης ψυχικής κόπωσης

Εξίσου σημαντικό είναι το υπολογιστικό κόστος του χρησιμοποιούμενου μοντέλου DL. Στην παρούσα εργασία δεν χρησιμοποιήθηκε GPU συμβατή με CUDA, με το μοντέλο TN να αναπτύσσεται στη CPU. Αυτός είναι ένας σημαντικός παράγοντας, δεδομένου ότι για μια συσκευή πραγματικού χρόνου, η παρουσία συμβατών με CUDA GPU είναι σπάνια. Με αυτό το δεδομένο, είναι ασφαλές να συμπεράνουμε ότι το μοντέλο ΤΝ θα αποδιδόταν εξίσου καλά σε έναν προηγμένο μικροεπεξεργαστή χωρίς την ανάγκη πυρήνων GPU. Αναλυτικότερα, η ανάλυση TFR για κάθε υποκείμενο ολοκληρώθηκε σε λιγότερο από 2 ms, ενώ ο χρόνος πρόβλεψης μιας μεμονωμένης περίπτωσης ήταν περίπου 4 ms, καθιστώντας το σύστημα κατάλληλο για ταξινομήσεις σε πραγματικό χρόνο και κατάλληλο για συστήματα EDGE-AI (όπου η εκπαίδευση και η εξαγωγή συμπερασμάτων πραγματοποιείται με ελάχιστη υπολογιστική ισχύ) [66]. Αντίθετα, οι εφαρμογές διαφορετικών αρχιτεκτονικών DL (όπως παρουσιάζονται στον Πίνακα 2) ή οι διαδικασίες υπολογισμού εισόδου οδηγούν σε αυξημένο (σε σύγκριση με το προτεινόμενο πλαίσιο) υπολογιστικό κόστος. Για παράδειγμα, ο υπολογισμός της PCA για την είσοδο απαιτεί περισσότερο από 10 φορές περισσότερο χρόνο σε σύγκριση με την TFR, ξεπερνώντας τα 20 ms για κάθε δείγμα [60]. Με βάση τα παραπάνω, η παρουσιαζόμενη προσέγγιση θα μπορούσε να θεωρηθεί ιδανική για την ανίχνευση κόπωσης σε πραγματικό χρόνο με πιθανές επεκτάσεις σε διεπαφές εγκεφάλου-υπολογιστή (BCI).

Εκτός από τις επιδόσεις του προτεινόμενου πλαισίου, ελέγχθηκαν τα χαρακτηριστικά που συνέβαλαν περισσότερο στη διαμόρφωση του μοντέλου ΤΝ. Εν προκειμένω, οι υπολογιζόμενες θετικές τιμές SHAP αναδεικνύουν την ισχύ του TFR στις ζώνες 5-15 Hz ως αυστηρό δείκτη ανάπτυξης νοητικής κόπωσης. Στην πραγματικότητα, προηγούμενες παρόμοιες μελέτες έχουν υποδείξει ότι οι αυξημένες απαιτήσεις εργασίας σε παραδείγματα μνήμης εργασίας μπορούν να αντικατοπτρίζονται ως νευρικές ταλαντώσεις θήτα (4-7 Hz) και άλφα (8-12 Hz) [67], [68]. Συγκεκριμένα, οι μεταβολές στην ισχύ του θήτα έχουν συσχετιστεί σταθερά με διαφορετικές καταστάσεις του επιπέδου κόπωσης, και ιδιαίτερα σε χειρισμούς μνήμης [48], [69]. Κατά παρόμοιο τρόπο, οι πτυχές της οπτικής προσοχής και της μνήμης εργασίας αναφέρουν συστηματικά μεταβολές της δραστηριότητας άλφα [70], [71]. Παρόλο που η κατάσταση κόπωσης προκλήθηκε ως αποτέλεσμα της στέρησης ύπνου και όχι της πειραματικής διαδικασίας, οι απαιτήσεις διατήρησης πληροφοριών και ανάκτησης στοιχείων του έργου n-back περιλαμβάνουν την ικανότητα της μνήμης εργασίας και τις υψηλές απαιτήσεις προσοχής [68]. Αυτό αναφέρεται επίσης στη δραστηριότητα της ζώνης βήτα (13-30 Hz) [72], [73], αν και η πολύ μικρή επικάλυψη των ευρημάτων μας με το εύρος συχνοτήτων βήτα δεν μπορεί να παράσχει μια πειστική υπόθεση. Όσον αφορά τις αρνητικές τιμές SHAP, τα ανασταλτικά στοιγεία που υποδεικνύονται στον χάρτη χρονοσυχνοτήτων 3-7 Hz υποδηλώνουν αλληλεπιδράσεις φορτίου μνήμης με κύματα θήτα, ενώ υποδηλώνουν αντιδραστικό έλεγχο που σχετίζεται με την εργασία (καθώς ο αντιδραστικός έλεγχος είναι λιγότερο απαιτητικός γνωστικά από τον προληπτικό έλεγχο) [74], [75].

Παρά το γεγονός ότι το προτεινόμενο πλαίσιο εμφάνισε συνολικά υψηλές επιδόσεις, η ερμηνεία των μεθόδων που εφαρμόστηκαν θα πρέπει να αντιμετωπίζεται με προσοχή. Ο κύριος προβληματισμός είναι ότι το υπό μελέτη σύνολο δεδομένων ενσωμάτωσε μικρό αριθμό μεμονωμένων παρατηρήσεων (20 άτομα), ο οποίος θεωρείται μέτριο δείγμα για σχεδιασμούς DL. Παρ' όλα αυτά, είναι παρόμοιο (ή μεγαλύτερο) από τις περισσότερες σχετικές παρόμοιες μελέτες [76]. Ένας άλλος πιθανός περιορισμός της παρούσας εργασίας είναι ο ορισμός της κατάστασης Fatigued και η ερμηνεία της ως βασική αλήθεια. Ωστόσο,

δεδομένου ότι τα δεδομένα προέρχονται από εφημερεύοντες γιατρούς και νοσηλευτικό προσωπικό, είναι ασφαλές να συμπεράνουμε ότι υπάρχει ψυχική κόπωση μετά τη βάρδια. Ως εκ τούτου, μπορεί να υποτεθεί ότι δεν υπάρχει προκατάληψη υποκειμενικότητας στην παρούσα έρευνα.

Λαμβάνοντας υπόψη τα αποτελέσματα της παρούσας μελέτης, στο μέλλον σκοπεύουμε να επεκτείνουμε τις συνολικές μεθοδολογίες, τόσο όσον αφορά τον πειραματικό όσο και τον μεθοδολογικό σχεδιασμό. Ως εκ τούτου, η ανάπτυξη προσεγγίσεων βαθιάς μάθησης για την ανίχνευση ψυχικής κόπωσης με χρήση σημάτων ΕΕG θα μπορούσε να ενισχυθεί με τη συμπερίληψη μεγαλύτερων συνόλων δεδομένων (που περιέχουν ποικιλία πληθυσμών και ρυθμίσεων, αυξάνοντας στη συνέχεια την ποικιλομορφία/μειώνοντας τις προκαταλήψεις του πληθυσμού), τα μέτρα γνωστικού φορτίου και τις παρατηρήσεις συμπεριφοράς, καθώς και τη βελτιστοποίηση στην αρχιτεκτονική DL. Αυτά, σε συνδυασμό με τις μεθόδους επεξηγηματικότητας, είναι πιθανό να αποφέρουν πολύτιμες γνώσεις και βελτιώσεις στην υποστρωμάτων που διέπουν την ψυχική κόπωση.

## 4. Κεφάλαιο 4 – Σύγχρονες τεχνικές εντοπισμού πηγής

Η τεχνική εντοπισμού πηγής (source localization – SL) από σήματα ΗΕΓ είναι μια απεικονιστική τεχνική που χρησιμοποιείται για τον εντοπισμό των ενεργοποιημένων περιοχών του εγκεφάλου, ενσωματώνοντας χρονικές και χωρικές συνιστώσες από σήματα ηλεκτροεγκεφαλογραφήματος (ΗΕΓ). Παρόλο που αυτή η απεικονιστική μέθοδος παρέχει υψηλή χρονική ανάλυση σε σύγκριση με άλλες απεικονιστικές μεθόδους, όπως η λειτουργική μαγνητική τομογραφία (fMRI), ο αριθμός των ηλεκτροδίων που χρησιμοποιούνται για την καταγραφή των ηλεκτρικών δυναμικών του εγκεφάλου είναι σημαντικά μικρότερος από τον αριθμό των ενεργοποιημένων νευρώνων, καθιστώντας τον εντοπισμό πηγών ΗΕΓ ένα κακώς τεθειμένο πρόβλημα (ill-posed problem). Για τον προσδιορισμό της πηγής γένεσης της ηλεκτρικής δραστηριότητας του εγκεφάλου, τα δυναμικά του τριχωτού της κεφαλής (δηλαδή, τα δυναμικά που δημιουργούνται από τη σύγχρονη ενεργοποίηση των πυραμιδικών νευρώνων στον εγκέφαλο [77], τα οποία διαδίδονται μέσω των διαφόρων ιστών που συνθέτουν την κεφαλή) καταγράφονται από μια συστοιχία επιφανειακών ηλεκτροδίων. Κατόπιν, μια εκτίμηση της προέλευσης των σημάτων υπολογίζεται μέσω της εξίσωσης Poisson [78]. Η εξίσωση Poisson προκύπτει άμεσα από τις εξισώσεις Maxwell, δεδομένου ότι οι ιστοί της κεφαλής έχουν τη διαπερατότητα του ελεύθερου χώρου, ενώ ο εντοπισμός των ενεργοποιημένων περιοχών εντός του εγκεφάλου αναφέρεται συχνά στη βιβλιογραφία ως το αντίστροφο πρόβλημα (inverse problem). Αντιθέτως, ο υπολογισμός των δυναμικών του τριγωτού της κεφαλής από μια δεδομένη πηγή (μια προαπαιτούμενη συνθήκη για τον υπολογισμό των αντίστροφων λύσεων) ορίζεται ως το ευθύ πρόβλημα του ΗΕΓ (EEG forward problem). Ως εκ τούτου, δεδομένου ότι η εξίσωση Poisson εξαρτάται σε μεγάλο βαθμό από τις ακριβείς τιμές των ηλεκτρικών και γεωμετρικών ιδιοτήτων (πάχος, αγωγιμότητα, κ.λπ.) των ιστών της κεφαλής (τριγωτό κεφαλής, κρανίο, εγκεφαλονωτιαίο υγρό (ENY), κοιλότητες εγκεφάλου, κ.λπ.), καθίσταται προφανές ότι ένα ακριβές μοντέλο κεφαλής είναι κρίσιμης σημασίας για την ορθή επίλυση του ευθέος προβλήματος. Για την αντιμετώπιση αυτού, ανατομικές πληροφορίες μπορούν να παρασχεθούν από τη μαγνητική τομογραφία (MT). Εντούτοις, είναι εξαιρετικά δύσκολο να ληφθεί MT κεφαλής με τρόπο εξατομικευμένο για κάθε υποκείμενο. Έτσι, ένας μεγάλος αριθμός μελετών βασίζεται σε πρότυπες ανατομικές πληροφορίες [79], [80], [81]. Η ύπαρξη ενός κατάλληλου μοντέλου κεφαλής διευκολύνει επίσης τη σωστή τοποθέτηση των ηλεκτροδίων, καθώς η θέση κάθε ηλεκτροδίου ποικίλλει για κάθε άτομο (αν και σε μικρό βαθμό). Σε αυτό το πλαίσιο, οι ανατομικές πληροφορίες καθώς και οι αντίστοιχες θέσεις των ηλεκτροδίων συνιστούν τον πίνακα πεδίου αγωγιμότητας (Leadfield matrix), ο ακριβής υπολογισμός του οποίου συνδέεται στενά με την ακρίβεια των λύσεων του αντίστροφου προβλήματος [82]. Ο εντοπισμός πηγών ΗΕΓ προβληματίζει τους ερευνητές, καθώς μια τεράστια ποικιλία συνδυασμών πηγών μπορεί να παράγει τα ίδια σήματα που καταγράφονται από το ΗΕΓ [83]. Αυτό αποτελεί θεμελιώδες πρόβλημα του εντοπισμού πηγών ΗΕΓ, διότι ακόμη και με ένα άψογο μοντέλο κεφαλής και ένα εξαιρετικό σήμα χωρίς θόρυβο, ο αριθμός των εξισώσεων που μπορούν να επιλυθούν είναι ελάχιστος σε σύγκριση με τον αριθμό των διπόλων (σημειακών πηγών) εντός του εγκεφάλου, οδηγώντας σε μη μοναδικές πιθανές λύσεις. Αυτή η μη-μοναδικότητα μπορεί να αντιμετωπιστεί με μαθηματικούς, ανατομικούς ή νευροφυσικούς περιορισμούς σχετικά με τη θέση και την κατεύθυνση των διπόλων [84], [85]. Επομένως, ακόμη και αν η μοναδικότητα δεν είναι εγγυημένη, αυτές οι εκ των προτέρων υποθέσεις (a priori assumptions) χρησιμοποιούνται προκειμένου να μειωθούν οι υπολογιστικές ανάγκες για την εκτίμηση της πηγής και να αυξηθεί η ακρίβεια των λύσεων.

Ωστόσο, χάρη στις πρόσφατες εξελίξεις που ενσωματώνουν καινοτόμες μεθοδολογίες, καθώς και την εισαγωγή προσεγγίσεων μηχανικής μάθησης στην επίλυση του αντίστροφου προβλήματος [86], [87], ο απαιτούμενος χρόνος και οι υπολογιστικοί πόροι για την επίλυση έχουν μειωθεί σημαντικά. Επιπλέον, εξελιγμένοι αλγόριθμοι έχουν μειώσει αποτελεσματικά το σφάλμα εντοπισμού, εκτιμώντας τη θέση και την ενεργοποίηση των διαφόρων φλοιϊκών περιοχών [88]. Αυτό είναι εμφανές σε αρκετές εφαρμογές, που κυμαίνονται από την ανίχνευση και την υποβοήθηση της θεραπείας της επιληψίας, έως την ερμηνεία συναισθημάτων από τις εκφράσεις του προσώπου [82], [89].

Αυτή η συνεχώς επικαιροποιούμενη πρόοδος τόσο στις λύσεις του ευθέος όσο και του αντίστροφου προβλήματος, καθιστά την SL ένα «ζωντανό» εργαλείο, που επιφέρει επανάσταση στη νευροδυναμική ανάλυση σε κλινικές, βιοϊατρικές και γνωσιακές εφαρμογές. Υπό αυτό το πρίσμα, η παρουσίαση των τρεχουσών εξελίξεων στον εντοπισμό πηγών ΗΕΓ αποτέλεσε το βασικό κίνητρο για την παρόν κεφάλαιο, το οποίο στοχεύει στην ανάδειξη των διαφόρων προσεγγίσεων των πολλαπλών αλγορίθμων SL αιχμής, επισημαίνοντας τα πλεονεκτήματα και τα μειονεκτήματα κάθε μεθόδου, καθώς και τους σχετικούς περιορισμούς τους που επηρεάζουν τη χωρική ανάλυση ή την υπολογιστική πολυπλοκότητα. Επιπλέον, συζητώνται οι πρόσφατες κλινικές και γνωσιακές εφαρμογές του εντοπισμού πηγών ΗΕΓ, καθώς και οι επιπτώσεις και οι προκλήσεις που ανακύπτουν λόγω μεθοδολογικών και τεχνικών περιορισμών, προτείνοντας μελλοντικές δράσεις.

## 4.1 Σύγχρονες μέθοδοι και προβλήματα

#### 4.1.1. Λύση εμπρόσθιου προβλήματος

Για την επίλυση του προβλήματος EEG προς τα εμπρός, απαιτούνται τρεις κύριες διαμορφώσεις: (α) τα μοντέλα κεφαλής και πηγής (δηλαδή η θέση των σημείων λύσης στον εγκέφαλο), (β) η ευθυγράμμιση των ηλεκτροδίων στο μοντέλο κεφαλής και (γ) ο πίνακας Leadfield χρησιμοποιώντας τις θέσεις των καναλιών σε σχέση με τις ανατομικές πληροφορίες του μοντέλου κεφαλής. Ως εκ τούτου, η ακρίβεια της λύσης εξαρτάται σε μεγάλο βαθμό από την αποτελεσματική παραγωγή και σύνθεση των ανωτέρω. Τα βασικά στοιχεία και η ροή των διαδικασιών παρουσιάζονται στο Σχήμα 4.1.

Οσον αφορά τη δημιουργία του μοντέλου κεφαλής, η ενσωμάτωση ανατομικών δεδομένων έχει καταστήσει παρωχημένα τα προηγούμενα, ευρέως υιοθετημένα σφαιρικά μοντέλα κεφαλής [90]. Η μετατροπή των εικόνων μαγνητικής τομογραφίας σε μοντέλα κεφαλής είναι μια χρονοβόρα διαδικασία με υψηλό υπολογιστικό κόστος και η διαδικασία αυτή παρουσιάζει βέλτιστα αποτελέσματα όσον αφορά την ακρίβεια εντοπισμού. Παρ' όλα αυτά, οι καταγραφές μαγνητικής τομογραφίας για το συγκεκριμένο υποκείμενο δεν είναι πάντα διαθέσιμες, με αποτέλεσμα να απαιτείται η κατασκευή του μοντέλου κεφαλής μέσω υπολογιστικών μεθόδων. Από αυτή την άποψη, ωστόσο, είναι σημαντικό να τονιστεί ότι η πολυπλοκότητα των μοντέλων κεφαλής δεν εμφανίζει γραμμική τάση με τον εντοπισμό του ΕSI, με αποτέλεσμα η υψηλή πολυπλοκότητα της υλοποίησης (π.χ. συμπερίληψη των σπονδυλωτών του κρανίου) να απαιτεί σημαντικούς υπολογιστικούς πόρους, με μικρή έως μηδαμινή επίπτωση στη χωρική ακρίβεια [91], [92].

Οι πιο συνηθισμένες μέθοδοι που χρησιμοποιούνται για τη μοντελοποίηση της κεφαλής είναι οι μέθοδοι οριακών στοιχείων (BEM), οι μέθοδοι πεπερασμένων διαφορών (FDM) και οι μέθοδοι πεπερασμένων στοιχείων (FEM) [12], [93], [94], [95]. Η προβολική μέθοδος (PM) (η οποία ενσωματώνει τους μαθηματικούς περιγραφείς των επιφανειών, όπως οι περιγραφείς Fourier, και τις μεθόδους μείωσης της διαστατικότητας, όπως η ανάλυση κύριων συνιστωσών), είναι επίσης καθιερωμένη, αν και λιγότερο διαδεδομένη [96], [97]. Σημειώνεται ότι για την PM, σε αντίθεση με τις προαναφερθείσες μεθόδους, ο χρόνος που απαιτείται για την εξώθηση του μοντέλου κεφαλής και οι υπολογιστικοί πόροι μειώνονται δραστικά, αλλά με χαμηλότερη χωρική ανάλυση [98].

Όσον αφορά τους συχνά χρησιμοποιούμενους αλγορίθμους, οι FEM και FDM παρέχουν πιο αναλυτικά αποτελέσματα και μπορούν να αντιμετωπίσουν προβλήματα ανομοιογένειας, γεγονός ζωτικής σημασίας εάν το μοντέλο κεφαλής απαιτεί τη μοντελοποίηση ανισοτροπικών ιδιοτήτων της λευκής ουσίας και του κρανίου. Από την άλλη πλευρά, η BEM παρουσιάζει χαμηλότερο υπολογιστικό χρόνο και ακρίβεια σε σύγκριση με τις FEM και FDM, μη μπορώντας να αντιμετωπίσει ανομοιογενή και μη γραμμικά προβλήματα.

Παρόλο που η προσέγγιση FEM παρέχει ένα πιο καθιερωμένο λεπτομερές μοντέλο, δεδομένου ότι ενσωματώνει περισσότερους από τους τρεις τυπικούς ιστούς της κεφαλής, η καθολική εφαρμογή της έχει περιοριστεί λόγω των υπολογιστικών αναγκών της και της έλλειψης εργαλείων ανοικτού κώδικα που ενσωματώνουν μια πλήρη προσέγγιση FEM. Ωστόσο, πρόσφατα εισήχθη ο αγωγός Fieldtrip-SimBio, ο οποίος παρέχει μια ολοκληρωμένη λύση του προωθητικού προβλήματος ΕΕG, ενώ χρησιμοποιεί ένα μοντέλο κεφαλής που έχει δημιουργηθεί με FEM [99]. Με βάση αυτή την παραδοχή, πρέπει επίσης να σημειωθεί ότι δεν είναι πάντα δυνατή η καταγραφή μαγνητικών τομογραφιών για κάθε υποκείμενο. Για το λόγο αυτό, δημιουργήθηκε ένα μέσο (τυποποιημένο) μοντέλο αγωγού όγκου με τη χρήση του ανατομικού προτύπου ICBM 152 και της FEM («The New York Head»), το οποίο έδειξε υποσχόμενα αποτελέσματα στην ESI, σε σύγκριση με άλλες τυποποιημένες προσεγγίσεις FEM και BEM [100]. Ενώ η αποτελεσματικότητα της FEM έχει αποδειχθεί ότι είναι συγκρίσιμη με τις αναλυτικές λύσεις [101], θα μπορούσαν να προκύψουν δυσμενή ζητήματα λόγω των φαινομένων διαρροής του κρανίου (δηλαδή της ανομοιογένειας του πάχους του κρανίου), τα οποία θα μπορούσαν να οδηγήσουν το υλοποιημένο μοντέλο να παρουσιάζει παρόμοιες ιδιότητες με ένα απλό μοντέλο σφαίρας τριών στρωμάτων [102]. Προκειμένου να ξεπεραστεί αυτή η ασυνέπεια, χρησιμοποιούνται διαφορετικές προσεγγίσεις FEM, οι οποίες αποκλίνουν από την τυπική συνεχή μέθοδο Galerkin-FEM (CG-FEM) δημιουργώντας μια μικτή ή ασυνεχή Galerkin-FEM που χρησιμοποιεί την προσέγγιση της αφαίρεσης [103], [104]. Τα στοιχεία δείχνουν ότι αυτή η λύση συνδυάζει τα οφέλη και από τις δύο προσεγγίσεις, μειώνοντας το φαινόμενο διαρροής κρανίου [19]. Επιπλέον, χρησιμοποιώντας αναλυτικές εκφράσεις σε συνδυασμό με την προσέγγιση αφαίρεσης, η ακρίβεια της λύσης του εμπρόσθιου προβλήματος αυξήθηκε σε σύγκριση με άλλες αριθμητικές προσεγγίσεις με παρόμοιο υπολογιστικό κόστος [105]. Αύξηση της ακρίβειας των αποτελεσμάτων του εμπρόσθιου προβλήματος ΕΕG έχει επίσης καταδειχθεί με την ενσωμάτωση της ανομοιογένειας των ιστών, ακόμη και εντός της φαιάς και της λευκής ουσίας [106]. Εκτός από τη FEM, υπολογιστικά αποδοτικές και ακριβείς λύσεις έχουν επίσης επιτευχθεί με την εισαγωγή εξισώσεων ανισοτροπικής αγωγιμότητας και τη χρήση του θεωρήματος της αμοιβαιότητας [107] σε μοντέλα κεφαλής FDM (AFDRM-NZ) [108]. Στο πλαίσιο αυτό, οι τιμές ανισοτροπικής αγωγιμότητας ενσωματώνονται με τη χρήση ενός συνόλου επιφανειακών ολοκληρωτικών εξισώσεων που ευθυγραμμίζουν τις παραγόμενες λύσεις με τα αναλυτικά αποτελέσματα [109]. Επισημαίνεται ότι η αγωγιμότητα του κρανίου και, ως εκ τούτου, η σωστή μοντελοποίηση είναι ένα κομβικό σημείο, το οποίο, λόγω ανωμαλιών της κεφαλής ή ανεπαρκών εργαλείων απεικόνισης, μπορεί να είναι δύσκολο να προσεγγιστεί [110]. Για την αντιμετώπιση αυτού του ζητήματος, οι προσεγγίσεις Bayesian Approximation Error (BAE) εμφάνισαν υψηλή αποτελεσματικότητα στη μείωση του σφάλματος εντοπισμού της πηγής κατά αρκετά χιλιοστά, ενισχύοντας τη χωρική ακρίβεια, η οποία είναι ζωτικής σημασίας στις κλινικές εφαρμογές [111].



Σχήμα 4.1 Ροή πληροφοριών και βασικά στοιχεία των εμπρόσθιων και αντίστροφων προβλημάτων.

#### 4.1.1 Αντίστροφο πρόβλημα

Σε αντίθεση με το εμπρόσθιο πρόβλημα, το αντίστροφο πρόβλημα δεν μπορεί να επιλυθεί μονοσήμαντα εάν δεν υπάρχουν εκ των προτέρων περιορισμοί σχετικά με τις θέσεις των πηγών [112]. Το γεγονός αυτό οδήγησε τους ερευνητές να ενσωματώσουν διάφορους μαθηματικούς περιορισμούς που προηγούνται της εκτίμησης της πηγής, μειώνοντας έτσι τον αριθμό των πιθανών λύσεων που προκύπτουν από τα καταγεγραμμένα δεδομένα. Τέτοιοι περιορισμοί αποτελούν τον πυρήνα πολλών από τις συμβατικές μεθόδους που χρησιμοποιούνται για την επίλυση του αντίστροφου προβλήματος. Οι αντίστροφες μέθοδοι που χρησιμοποιούνται συνήθως εμπίπτουν σε δύο κύριες κατηγορίες - τη μη παραμετρική (μη προσαρμοστική) και την παραμετρική (προσαρμοστική) απεικόνιση κατανεμημένης πηγής.

Αν και υπάρχει ένας εξαιρετικά μεγάλος αριθμός μεθοδολογιών ESI για παραμετρικές (π.χ. Beamformers, Multiple Signal Classification) και μη παραμετρικές (π.χ. Minimum Norm Estimate, Focal Underdetermined System Solution, Local Auto-Regressive Average) μεθόδους κατανεμημένης απεικόνισης πηγών της αντίστροφης λύσης, στην παρούσα εργασία, εστιάζουμε στους πιο συχνά χρησιμοποιούμενους αλγορίθμους, για να αναδείξουμε την πρόοδο της ESI. Για μια εκτενή ανασκόπηση των μαθηματικών πτυχών και των ιδιοτήτων αρκετών από αυτές τις συμβατικές μεθόδους και τις παραλλαγές τους, που ξεφεύγουν από το πεδίο εφαρμογής αυτής της εργασίας ανασκόπησης, παρακαλούμε ανατρέξτε στις διεξοδικές ανασκοπήσεις [10], [113].

Οι πιο συχνοί μη παραμετρικοί αλγόριθμοι που χρησιμοποιούνται είναι η λύση της εκτίμησης της ελάχιστης νόρμας (MNE) [114] και η σταθμισμένη κατά βάθος παραλλαγή της (dw-MNE) [115], [116], αν και αρκετοί άλλοι σχεδιασμοί λαμβάνουν υπόψη τις ίδιες αρχές με τροποποιημένες ρυθμίσεις και ενσωμάτωση πρόσθετων παραμέτρων. Για παράδειγμα, η δραστηριότητα ηλεκτρομαγνητικής τομογραφίας χαμηλής ανάλυσης (LORETA) εκτιμά την πυκνότητα ρεύματος που δίνεται από τη λύση της ελάχιστης νόρμας, αλλά με μια πιο εξελιγμένη κανονικοποίηση, χρησιμοποιώντας έναν διακριτό τελεστή Laplace που επιλέγει κατά προτίμηση κατανομές πηγών που διαχέονται («ομαλές»), σε αντίθεση με τον πίνακα ταυτότητας της MNE [18], [117]. Από την άλλη πλευρά, οι παραμετρικές μέθοδοι απεικόνισης κατανεμημένης πηγής περιλαμβάνουν συνήθως διαμορφωτές δέσμης με γραμμική περιορισμένη ελάχιστη διακύμανση (LCMV) που εξαρτώνται από δομικά συναφή φίλτρα για να παρέχουν αποτελεσματικό εντοπισμό πηγής ανεξάρτητα από τη συνδιακύμανση του θορύβου [118]. Σε αντίθεση με τις μη παραμετρικές προσεγγίσεις για την επίλυση της αντίστροφης λύσης, οι διαμορφωτές δέσμης LCMV απομονώνουν τα σήματα που παράγονται από διαφορετικά μέρη του εγκεφάλου χρησιμοποιώντας χωρικά φίλτρα, επιτρέποντας έτσι τους υπολογισμούς της λύσης να γίνονται ανεξάρτητα για κάθε σημείο λύσης.

Παρόλο που οι συμβατικές μέθοδοι έχουν αποδειχθεί αποτελεσματικές για τον προσδιορισμό των ενεργοποιημένων περιοχών του εγκεφάλου, οι αναδυόμενες τεγνολογίες στους τομείς της μηχανικής και της βαθιάς μάθησης εισήχθησαν πρόσφατα στο ESI. Ως εκ τούτου, μια νέα μέθοδος που προτάθηκε για την επίλυση του αντίστροφου προβλήματος χρησιμοποιεί την αρχιτεκτονική βαθιών αναδρομικών νευρωνικών δικτύων με μονάδες μακράς βραχυπρόθεσμης μνήμης (LSTM) σε ένα πλαίσιο αυτόματου κωδικοποιητή, παρουσιάζοντας εξαιρετικό μέσο σφάλμα εντοπισμού μικρότερο από 5 mm σε προσομοιωμένα δεδομένα μίας πηγής [86]. Αυτή η αρχιτεκτονική είναι σε θέση να μοντελοποιήσει τις χωροχρονικές πληροφορίες που παρέχονται με την εκπαίδευση του δικτύου ώστε να αντιλαμβάνεται τη συσγέτιση μεταξύ της θέσης της πηγής και των σημάτων ΗΕΓ χωρίς να γρειάζονται εκ των προτέρων περιορισμοί, που συνήθως παρέχονται χειροκίνητα από πιο συμβατικές μεθόδους. Με παρόμοιο τρόπο, το ConvDip, ένα συνελικτικό νευρωνικό δίκτυο (CNN), έχει επιδείξει χαμηλότερο κανονικοποιημένο μέσο τετραγωνικό σφάλμα στις λύσεις ESI σε σύγκριση με αυτό της ακριβούς LORETA (eLORETA) και των διαμορφωτών δέσμης για μία μόνο πηγή, χρησιμοποιώντας ένα ρηχό CNN με ένα συνελικτικό στρώμα και δύο πλήρως συνδεδεμένα στρώματα [119]. Σε σύγκριση με την προσέγγιση LSTM, το ConvDip εκπαιδεύτηκε με μεμονωμένες χρονικές περιπτώσεις σε προσομοιωμένα δεδομένα αλλά με πολλαπλές πηγές. Αυτό είναι σημαντικό, καθώς μια μεγάλη ποικιλία αντίστροφων λύσεων, όπως η eLORETA και ο LCMV beamformer, βασίζονται σε πίνακες συνδιακύμανσης θορύβου που υπολογίζονται με τη χρονική πληροφορία των σημάτων EEG, επηρεάζοντας σημαντικά την ακρίβεια του μοντέλου εάν αυξηθεί ο θόρυβος. Επιπλέον, το γεγονός ότι τα απλά δίκτυα μπορούν να μάθουν τα πρότυπα μεμονωμένων χρονικών σημείων και να προβλέψουν λογικές αντίστροφες λύσεις είναι ένα σημαντικό σημείο για τη μείωση της πολυπλοκότητας και κατά συνέπεια του υπολογιστικού κόστους. Πρόσθετα πλαίσια νευρωνικών δικτύων ESI περιλαμβάνουν έναν αυτόματο κωδικοποιητή αποθορυβοποίησης (DST-DAE) που αποτελείται από έξι στρώματα, τρία μπλοκ κωδικοποίησης και τρία μπλοκ αποκωδικοποίησης. Αυτή η μέθοδος ήταν σε θέση να χαρτογραφήσει απευθείας τα σήματα του EEG και του μαγνητοεγκεφαλογραφήματος (MEG) στις φλοιώδεις πηγές, μειώνοντας το σφάλμα εντοπισμού σε λιγότερο από ένα γιλιοστό [120]. Εν προκειμένω, αξιοποιούνται τόσο οι χρονικές όσο και οι χωρικές πληροφορίες σε συνθετικά δεδομένα, προκειμένου να εξαχθεί η αντίστροφη χαρτογράφηση. Το κύριο πλεονέκτημα αυτής της μεθόδου έγκειται στην ανθεκτικότητα και την ευρωστία της έναντι χαμηλού λόγου σήματος προς θόρυβο (SNR), με αποτέλεσμα την αποτελεσματική εκτίμηση της πηγής με εξαιρετικές ιδιότητες αποθορυβοποίησης. Αν προστεθούν όλες μαζί, οι πρόσφατες εξελίξεις στην εκτίμηση ESI με μηχανική μάθηση υποδηλώνουν την αποτελεσματικότητα των χρησιμοποιούμενων διαδικασιών σε αντίθεση με τις παραδοσιακές προσεγγίσεις που βασίζονται σε μοντέλα, ιδίως δεδομένου ότι σε όλες τις μεθόδους που βασίζονται σε δεδομένα, χρησιμοποιήθηκαν πολύ λίγες ή καθόλου μαθηματικές προκαταλήψεις, ενώ απουσιάζει η ανάγκη βελτιστοποίησης των παραμέτρων για νέα δεδομένα.

Εκτός από τα σχέδια βαθιάς μάθησης, πρόσφατες μελέτες περιλαμβάνουν ιεραρχικές μεθόδους ανάλυσης Bayesian για τον εντοπισμό πηγών. Η σημασία των μοντέλων Bayesian βασίζεται στην ενσωμάτωση στατιστικών εκ των προτέρων πληροφοριών σχετικά με τις πηγές, εξαλείφοντας κοινά προβλήματα όπως οι πηγές-φαντάσματα και η ασυσγέτιστη μετάβαση ενεργοποίησης [10]. Το κύριο πλεονέκτημα αυτών των λύσεων είναι ο συνδυασμός της μάθησης με βάση τα δεδομένα με αραιά priors, ελαχιστοποιώντας τη συνάρτηση κόστους και μεγιστοποιώντας την πιθανότητα συσχετισμένων πηγών [121]. Πιο πρόσφατα, εισήχθη μια τροποποιημένη προσέγγιση Bayes, εφαρμόζοντας τη μικτή νόρμα <sub>ε20</sub> αντί της <sub>ε21</sub> (που χρησιμοποιείται κυρίως με τις Bayesian Methods) και πολυμεταβλητές Bernoulli Laplacian priors [88], με την κύρια διαφορά μεταξύ των Bayesian Models να είναι οι κατανομές πιθανότητας για τη συσγέτιση των πηγών. Η μέθοδος αυτή ήταν σε θέση να παρέχει πιο αραιές λύσεις, ελαχιστοποιώντας την υποεκτίμηση της έντασης των ενεργοποιήσεων, υποδεικνύοντας υψηλότερη απόδοση εντοπισμού από άλλες συμβατικές μεθόδους Bayesian, τόσο σε προσομοιώσεις όσο και σε πραγματικά ακουστικά και οπτικά προκλητά δεδομένα, αλλά με υψηλό υπολογιστικό κόστος, απαιτώντας σχεδόν 58 φορές περισσότερο χρόνο από τις προσεγγίσεις με μικτή νόρμα (21. Σε μια σχετική μελέτη [122], ένας υπολογιστικά αποδοτικός αλγόριθμος Expectation-Maximization με τη χρήση φίλτρου Kalman σταθερής κατάστασης (SS-KF) και σταθερού σταθερού διαστήματος εξομάλυνσης (SS-FIS) παρείχε σημαντική βελτίωση της απόδοσης σε σύγκριση με άλλες υπάρχουσες μεθόδους, ενώ ενθυλάκωνε τις χωρικές εξαρτήσεις μεταξύ των πηγών. Ο μετριασμός του υπολογιστικού φόρτου βασίζεται στη χρήση των SS-KF και SS-FIS που υπολογίζονται μία φορά καθ' όλη τη διάρκεια της εκτίμησης των πηγών χωρίς απώλεια της ακρίβειας του μοντέλου, μειώνοντας στη συνέχεια 12 φορές το χρόνο που απαιτείται για την εξαγωγή των αποτελεσμάτων σε σύγκριση με την πλήρη μέθοδο KF/FIS. Θα πρέπει να επισημανθεί ότι οι προαναφερθείσες μέθοδοι παρουσιάζουν θεωρητικά μηδενικό σφάλμα εντοπισμού, ξεπερνώντας τις λύσεις που παρέχουν τα νευρωνικά δίκτυα. Ωστόσο, το πλεονέκτημα των νευρωνικών δικτύων έγκειται στον μικρό χρόνο που απαιτείται για τον υπολογισμό της αντίστροφης λύσης. Από την άποψη αυτή, η δυνατότητα γρήγορου υπολογισμού των εκτιμήσεων της πηγής είναι υψίστης σημασίας στη φιλοδοξία ανάπτυξης εφαρμογών ESI πραγματικού χρόνου στο εγγύς μέλλον.

## 4.2 Εφαρμογές ESI-κλινικής και γνωστικής έρευνας

Ο εντοπισμός της πηγής του εγκεφάλου παρέχει μια νέα διάσταση των καταγεγραμμένων δεδομένων ΗΕΓ, η οποία μπορεί να εφαρμοστεί σε διάφορες συνθήκες, τόσο σε κλινικές όσο και σε γνωστικές εφαρμογές. Σημειωτέον ότι νέοι αλγόριθμοι βελτιώνουν συνεχώς τον χρόνο/αποτελεσματικότητα και την ακρίβεια των λύσεων ΕSI. Για το λόγο αυτό, ο αριθμός των εφαρμογών στις οποίες εφαρμόζεται το ESI αυξάνεται συνεχώς. Κατά συνέπεια, σε αυτή την εργασία ανασκόπησης ήταν αδύνατο να συμπεριληφθούν όλες οι νέες εφαρμογές της απεικόνισης πηγής. Στόχος μας ήταν να προσπαθήσουμε να καταδείξουμε τις προόδους στον τομέα παρουσιάζοντας τα πιο ενδεικτικά παραδείγματα, καταδεικνύοντας τη σημασία της ESI για κλινικές και γνωστικές εφαρμογές.

Κλινικές εφαρμογές
Η σημαντικότερη απασχόληση της ESI είναι σε κλινικές καταστάσεις. Η υψηλή χρονική ανάλυση που παρέχει το HEΓ, σε συνδυασμό με τη χωρική εκτίμηση των πηγών που δίνει το ESI, παρέχει μη επεμβατική παρακολούθηση συγκεκριμένων περιοχών του εγκεφάλου που δεν μπορεί να προσφέρει η ανάλυση με μία μόνο μέθοδο. Ως εκ τούτου, το ESI χρησιμοποιείται συνήθως για τον εντοπισμό των επιληπτικών εστιών, ζωτικής σημασίας για τη θεραπεία της επιληψίας και/ή για την επακόλουθη χειρουργική επέμβαση [123]. Σε μια πρόσφατη μακροχρόνια μελέτη [124], μια σύγχρονη καταγραφή υψηλής πυκνότητας HEΓ και MEG για τον εντοπισμό της πηγής της ικτικής και διαικτικής δραστηριότητας παρείχε ακριβέστερα αποτελέσματα όταν η ανάλυση διεξήχθη μόνο για κάθε μια από τις δύο μορφές από ό,τι η πολυτροπική προσέγγιση. Προτείνεται επίσης να δοθεί λιγότερη έμφαση στην ανάλυση της ικτικής εκφόρτισης στη μέση φάση και να συμπεριληφθεί η φάση της όψιμης αιχμής στις λύσεις πηγής. Ο εντοπισμός της πηγής του ΗΕΓ ως τεχνική μίας μονής διατροπής είναι επίσης μια αποτελεσματική τεχνική εντοπισμού για την εστιακή επιληψία, ελαχιστοποιώντας το κόστος της ταυτόχρονης καταγραφής διαφορετικών διατροπών, που απαιτεί πρόσθετο ακριβό εξοπλισμό [125], [126].

Πέρα από τα προφανή οικονομικά και επιστημονικά οφέλη των εφαρμογών ESI, η μη επεμβατικότητα του εντοπισμού πηγής παρουσιάζει σημαντικά κοινωνικά και υγειονομικά πλεονεκτήματα, δεδομένου ότι οι (πολλαπλές) ενδοκρανιακές επεμβατικές καταγραφές μπορούν να αποφευχθούν σε ορισμένους τύπους επιληψίας, ανακουφίζοντας έτσι τους ασθενείς από το προεγχειρητικό άγχος, καθώς και από τον κίνδυνο λοιμώξεων της χειρουργικής περιοχής. Για το λόγο αυτό, η επικρατέστερη χρήση του εντοπισμού της πηγής του ΗΕΓ είναι η προεγχειρητική αξιολόγηση επιληπτικών ασθενών με φαρμακευτική αγωγή [127]. Κατά τη διάρκεια αυτής της διαδικασίας, οι πληροφορίες που παρέχει το ΗΕΓ σχετικά με τις επιληπτικές εκφορτίσεις μπορούν να έχουν σημαντικό αντίκτυπο στο σχέδιο διαχείρισης, όσον αφορά την ακριβέστερη τοποθέτηση των ενδοκρανιακών ηλεκτροδίων [128], [129]. Ο προσδιορισμός των ζωνών έναρξης των επιληπτικών κρίσεων μπορεί όχι μόνο να καθορίσει τις ακριβείς φλοιώδεις δομές για χειρουργική εξαγωγή, αλλά και να παράσχει σημαντικές πληροφορίες σχετικά με τη νευροδυναμική κατάσταση της επιληπτικής εγκεφαλικής δραστηριότητας, ιδίως σε γενικευμένες (π.χ. επιληψία απουσίας) [130] ή εστιακές επιληψίες (π.χ. δομική, μη λυτική επιληψία) [52,57,58].

Αν και το ΗΕΓ υψηλής πυκνότητας δεν είναι πάντα διαθέσιμο, ο εντοπισμός πηγής ΗΕΓ χαμηλής πυκνότητας ήταν σταθερά αποτελεσματικός στον εντοπισμό της επιληπτικής εγκεφαλικής δραστηριότητας, επιτυγχάνοντας ακρίβεια άνω του 90% σε ορισμένες περιπτώσεις, ενώ υποδεικνύει ότι η απόδοση θα μπορούσε να βελτιωθεί περαιτέρω με τη συμπερίληψη της ανάλυσης συνδεσιμότητας [131], [132], [133]. Η ESI έχει επίσης χρησιμοποιηθεί στην παιδιατρική επιληψία. Σε μια πρόσφατη μελέτη [134], χρησιμοποιώντας μια οπτική εργασία μνήμης εργασίας ως ερέθισμα και ένα πλαίσιο εντοπισμού πηγής, εντοπίστηκαν κρίσιμοι βιοδείκτες (που δεν μπορούσαν να εντοπιστούν μόνο με την ανάλυση σε επίπεδο αισθητήρων), οδηγώντας σε μια πιο εντοπισμένη και καθορισμένη θεραπευτική πολιτική.

Η λεπτή χρονική επικοινωνία των διαφόρων περιοχών του εγκεφάλου στο χώρο των πηγών έχει επίσης εφαρμοστεί σε νευροεκφυλιστικές ασθένειες [135]. Συχνά, χρησιμοποιείται στην αξιολόγηση της εξέλιξης της νόσου του Alzheimer ή ακόμη και στην πρόγνωσή της, δεδομένου ότι η εγκεφαλική ενεργοποίηση που σχετίζεται με τις γνωστικές επιδόσεις μεταβάλλεται συνήθως κατά την εξέλιξη της νόσου [136]. Ως εκ τούτου, η eLORETA και η

ανάλυση συνδεσιμότητας έχουν υιοθετηθεί για τη διάγνωση της νόσου του Alzheimer [137], [138]. Μια άλλη ενδιαφέρουσα προσέγγιση είναι η χρήση της ESI στον εντοπισμό της προέλευσης των επιληπτικών αιγμών γρησιμοποιώντας τις διαφορές που δημιουργούνται από τον ύπνο μη ταχείας κίνησης των ματιών (NREM) και ταχείας κίνησης των ματιών (REM) [139]. Λαμβάνοντας υπόψη το γεγονός ότι το ΗΕΓ έχει χρησιμοποιηθεί σταθερά για τη μελέτη των κύκλων και των διαταραχών που σχετίζονται με τον ύπνο, ο εντοπισμός της πηγής μπορεί να προσδιορίσει αποτελεσματικά τις κύριες περιογές που εμπλέκονται στη μετάβαση από την έναρξη του ύπνου και τις μεταβολές της δραστηριότητας στις περιοχές του εγκεφάλου που επηρεάζονται από την αποκατάσταση της στέρησης ύπνου [140], [141]. Επιπλέον, η ESI είναι αποτελεσματική στην ανίχνευση των περιοχών και της σχετικής δραστηριότητας κατά τη διάρκεια διαφορετικών σταδίων ύπνου NREM με τη χρήση καταγραφών MEG και EEG [142], των περιοχών που είναι υπεύθυνες για τη δημιουργία ατράκτων ύπνου [143] και τη δημιουργία αργών κυμάτων [144]. Είναι ενδιαφέρον ότι ο εντοπισμός της πηγής έγει εφαρμοστεί για την αποκωδικοποίηση της προέλευσης της υπνοβασίας και του τρόμου του ύπνου, που χρησιμοποιήθηκε όχι μόνο για την εκτίμηση των εγκεφαλικών περιοχών που επηρεάστηκαν από τη διαταραχή διέγερσης του ύπνου, αλλά και για τον υπολογισμό της μείωσης της ισχύος της δραστηριότητας των αργών κυμάτων [145].

Παρά τις προγνωστικές και διαγνωστικές δυνατότητες της ESI, οι διαδικασίες αποκατάστασης μπορούν επίσης να υποστηριχθούν ουσιαστικά. Σε αυτό το πλαίσιο, η διάκριση της κατεύθυνσης μέσω της αποκωδικοποίησης των προθέσεων κίνησης σε άτομα που πάσχουν από κινητικά ελλείμματα (όπως οι ασθενείς με εγκεφαλικό επεισόδιο) είναι υψίστης σημασίας για τη βέλτιστη διαχείριση και αξιολόγηση του σχεδιασμού αποκατάστασης. Πολλαπλές προσεγγίσεις έχουν συμβάλει στην ταξινόμηση της κίνησης του χεριού με πολλαπλές κατευθύνσεις, χρησιμοποιώντας σταθμισμένο εντοπισμό πηγής ΜΝΕ [146] ή σταθμισμένο MNE με επικάλυψη (Overlapping Averaging weighted MNE, OAwMNE) [147]. Επιπλέον, η ανίχνευση των θέσεων της εγκεφαλικής πηγής σύνθετων φανταστικών κινήσεων, όπως η κάμψη, η έκταση, η υπερέκταση και η πρηνισμός, είναι επίσης εφικτή με μεθόδους εντοπισμού πηγής και αλγορίθμους επιλογής χαρακτηριστικών [75,76]. Με αυτή την προϋπόθεση, η ESI έχει χρησιμοποιηθεί για την αξιολόγηση της κατανομής των εγκεφαλικών λειτουργιών όσον αφορά τις προσπάθειες που κάνουν οι ασθενείς με εγκεφαλικό επεισόδιο στην κίνηση των άκρων [150]. Πρόσφατες προσεγγίσεις έχουν χρησιμοποιήσει τεχνικές ΗΕΓ υψηλής πυκνότητας και beamformer για να διευκολύνουν την ανάλυση της λειτουργικής συνδεσιμότητας, μετρώντας έτσι τη σύνδεση μεταξύ της συνοχής των ζωνών άλφα και βήτα με την κινητική μάθηση και τη φλοιώδη πλαστικότητα των διαταραχών μετά το εγκεφαλικό επεισόδιο [151], [152], [153].

#### Γνωστικές εφαρμογές

Εκτός από τις κλινικές εφαρμογές, ο εντοπισμός της εγκεφαλικής πηγής βοηθά στη σταδιακή διαλεύκανση των υποκείμενων νευρικών υποστρωμάτων σε διάφορες καταστάσεις και ψυχικές καταστάσεις. Ως εκ τούτου, η ESI έχει χρησιμοποιηθεί με συνέπεια στην αποκρυπτογράφηση των γνωστικών λειτουργιών σε διάφορες εργασίες, όπως η ερμηνεία λέξεων και η ανάλυση κόπωσης.

Πολλαπλά παραδείγματα περιλαμβάνουν τον εντοπισμό της πηγής στα δυναμικά που σχετίζονται με γεγονότα (ERPs) για να φωτίσουν την εγκεφαλική δραστηριότητα που σχετίζεται με ορισμένα ερεθίσματα ή γεγονότα. Για παράδειγμα, οι αποκρίσεις που

σχετίζονται με γεγονότα σε καθήκοντα Go-Nogo σε συνδυασμό με ESI έχουν χρησιμοποιηθεί για τον εντοπισμό και την ερμηνεία της αναστολής της σωματοαισθητικής απόκρισης [154], [155]. Συγκεκριμένα, η ανάλυση ERP και η τυποποιημένη LORETA (sLORETA) χρησιμοποιήθηκαν για να συνδέσουν τις διαφορές του πλάτους της συνιστώσας Nogo-N2 ERP με τις αλλαγές της ενεργοποίησης στη δεξιά κάτω έλικα, δείχνοντας ότι η αναστολή της απόκρισης συσχετίζεται με τη διαμόρφωση του αισθητηριακού κέρδους [155]. Μια άλλη πρόσφατη μελέτη [156] χρησιμοποίησε beamformers στην ανάλυση ERP για να παράσχει στοιχεία για τις περιοχές που είναι υπεύθυνες για τη δημιουργία του σήματος P300 που σχετίζονται με τις λειτουργίες προσοχής και μνήμης του εγκεφάλου.

Η ερμηνεία των εκφράσεων του προσώπου και των συναισθημάτων στο χώρο πηγής έχει επίσης διερευνηθεί με συνέπεια σε διάφορες μελέτες. Ως εκ τούτου, λαμβάνοντας υπόψη ότι η γνωστική επεξεργασία του προσώπου εξελίσσεται κατά τα πρώτα στάδια της ζωής, η απασχόληση της LORETA μπόρεσε να παράσχει εικόνα της υποκείμενης διαδικασίας [157]. Με παρόμοιο τρόπο, χρησιμοποιώντας έναν αλγόριθμο MNE, η ESI μπόρεσε να προωθήσει περαιτέρω την αποκωδικοποίηση της κατάστασης του εγκεφάλου αποκρυπτογραφώντας τις νευρωνικές λειτουργίες της αναγνώρισης προσώπου [158], ενώ η αξιοποίηση της μεθόδου LAURA και η αξιοποίηση του «φαινομένου Kuleshov» ανέδειξε αποτελεσματικά τις γνωστικές λειτουργίες της έκφρασης του προσώπου [159]. Επιπλέον, ο συνδυασμός της ESI με έναν αυτο-οργανωτικό χάρτη χαρακτηριστικών (SOMF) είχε ως αποτέλεσμα την ακριβή κατηγοριοποίηση των συναισθημάτων που γίνονται αντιληπτά από φυσιογνωμικά στοιχεία, (έως και 91% ακρίβεια ταξινόμησης) [89]. Παρόμοιες μελέτες εντοπισμού πηγών έχουν επίσης διεξαχθεί για τον εντοπισμό και την ανάλυση της επεξεργασίας των χαρακτηριστικών του προσώπου στον αυτισμό, διακρίνοντας τις περιοχές που ενεργοποιούνται σε διαφορετικές περιπτώσεις [160].

Πιο πρόσφατα, η ESI έχει εφαρμοστεί σε οπτικά και ακουστικά προκλητά δεδομένα, σε μια προσπάθεια κατανόησης των υποκείμενων εγκεφαλικών συνδέσεων και των σχετικών ανασταλτικών δικτύων [81], [161]. Στην πραγματικότητα, δεδομένου ότι η απεικόνιση πηγής είναι σε θέση να αναλύσει τις διαδρομές της εγκεφαλικής δραστηριότητας μέσω του χρόνου, η ανάλυση συνδεσιμότητας παρουσιάζει μια νέα προοπτική για τον εντοπισμό της μετάδοσης πληροφοριών μεταξύ των περιοχών του φλοιού. Στο πλαίσιο αυτό, η κατανόηση της γλώσσας και ο αντίκτυπος στον εγκέφαλο έχει μελετηθεί με τη χρήση της συνδεσιμότητας και της sLORETA, αποκαλύπτοντας μεμονωμένα δίκτυα που δραστηριοποιούνται κατά τη λεξικοσημασιολογική επεξεργασία μεμονωμένων λέξεων [162]. Επιπλέον, πλαίσια αναγνώρισης προτύπων σε συνδυασμό με ESI έχουν εντοπίσει αποτελεσματικά κοινές και διακριτές λειτουργικές συνδέσεις (που αντιπροσωπεύουν μια σύνδεση μεταξύ των διαφόρων δομών του εγκεφάλου) σε διαφορετικά επίπεδα φόρτου εργασίας σε δύο εργασίες μνήμης εργασίας [163]. Επιπλέον, σε πειράματα προσομοίωσης 2D και 3D προσομοίωσης πτήσης, η eLORETA και η θεωρία γραφημάτων χρησιμοποιήθηκαν για την αξιολόγηση της δυναμικής επανασύνδεσης των λειτουργικών συνδέσεων μεταξύ των περιογών του εγκεφάλου ως αποτέλεσμα του πνευματικού φόρτου εργασίας, δημιουργώντας τη βάση για πιθανές εφαρμογές στον πραγματικό κόσμο [164].

#### 4. Προκλήσεις και περιορισμοί του ESI

Κατά την τελευταία δεκαετία έχουν σημειωθεί σημαντικές πρόοδοι όσον αφορά τις μεθοδολογίες ESI, τόσο στον υπολογιστικό χρόνο όσο και στην ακρίβεια των λύσεων για το

αντίστροφο πρόβλημα του ΗΕΓ. Ωστόσο, η ίδια η φύση του αντίστροφου προβλήματος, δεδομένων των υποκαθορισμένων ιδιοτήτων του, αποτελεί σημαντικό εμπόδιο. Παρά το γεγονός ότι οι πρόσφατες προσεγγίσεις έχουν μετριάσει επιτυχώς αρκετούς περιορισμούς, εξακολουθούν να υπάρχουν ορισμένες προκλήσεις που πρέπει να αντιμετωπιστούν.

Το μεγαλύτερο μειονέκτημα του ΗΕΓ, και κατ' επέκταση των υπολογισμών εντοπισμού πηγής, είναι η ευαισθησία του στο θόρυβο του σήματος, ο οποίος μπορεί να αποτελείται από θόρυβο υποβάθρου, παραμορφώσεις και τεχνουργήματα κίνησης. Αυτές οι παρεμβολές επηρεάζουν σε μεγάλο βαθμό την ακρίβεια της αντίστροφης λύσης, παράγοντας πηγέςφαντάσματα ή ακόμη και μετατοπίζοντας εντελώς τις προβλεπόμενες περιοχές εγκεφαλικής δραστηριότητας [165]. Στο πλαίσιο αυτό, τα αντίστοιχα αποτελέσματα μπορεί να είναι παραπλανητικά και ενδέχεται να παρουσιάζουν κινδύνους για την ασφάλεια, ιδίως όσον αφορά τον προεγχειρητικό σχεδιασμό και την αξιολόγηση. Για την αντιμετώπιση αυτών των προβλημάτων έχουν προταθεί διάφορες μέθοδοι ειδικής επεξεργασίας σήματος, όπως η Ανάλυση ανεξάρτητων συνιστωσών [166] για την αφαίρεση των τεχνουργημάτων και οι καταγραφές Faraday/κενών δωματίων, προκειμένου να επιλυθεί ο θόρυβος υποβάθρου και οργάνων. Ωστόσο, μπορεί να εξακολουθεί να υπάρχει άσχετος θόρυβος σήματος από ανασταλτικές εγκεφαλικές λειτουργίες, ενώ οι κρίσιμες ιδιότητες του σήματος θα μπορούσαν να ενσωματωθούν στις αναγνωρισμένες ανεξάρτητες συνιστώσες που σχετίζονται με τα τεχνουργήματα και συνεπώς να αφαιρεθούν μαζί με αυτές [167], [168].

Ο αριθμός των ηλεκτροδίων και ο αριθμός των διπόλων που ορίζονται από τη δημιουργία του μοντέλου πηγής είναι επίσης κρίσιμος. Όπως περιγράφηκε στις προηγούμενες ενότητες, ο αριθμός των σημείων πηγής υπερβαίνει κατά πολύ τον αριθμό των ηλεκτροδίων και, ως εκ τούτου, η λύση δεν είναι ποτέ μοναδική και μπορεί να εκτιμηθεί μόνο με τη χρήση προτεραιοτήτων (χωρικών, μαθηματικών ή ανατομικών). Επιπλέον, ένας πολύ μεγάλος αριθμός σημείων πηγής θα αύξανε την ανάλυση της λύσης ESI, αλλά θα αύξανε επίσης την πολυπλοκότητα της διαδικασίας. Πιο συγκεκριμένα, ένας μικρός αριθμός διπόλων θα οδηγούσε σε χαμηλότερη χωρική ανάλυση και ακρίβεια θέσης, καθιστώντας προφανώς την παραγόμενη λύση πιο επιρρεπή σε παρερμηνείες, αλλά απαιτώντας λιγότερη μνήμη και υπολογιστικό χρόνο για τον υπολογισμό της. Από την άλλη πλευρά, ένας μεγαλύτερος αριθμός σημείων πηγής οδηγεί σε λύσεις με υψηλότερη χωρική ανάλυση με μικρότερη αραιότητα της εγκεφαλικής δραστηριότητας γύρω από κάθε υπολογισμένο σημείο, λόγω του μεγαλύτερου αριθμού γειτονικών σημείων [167]. Ο κύριος λόγος για την αυξημένη ακρίβεια είναι ότι ο πίνακας Leadfield είναι πληρέστερος και υπολογίζεται διεξοδικότερα, απαιτώντας αναμφίβολα πρόσθετους πόρους. Ως εκ τούτου, υπάρχει μια λεπτή ισορροπία σχετικά με το υπολογιστικό κόστος και την έξοδο ESI, όπου πρέπει να εξεταστεί η ελαχιστοποίηση της αραιότητας της φλοιώδους δραστηριότητας (δηλαδή, η έκταση της ενεργοποίησης του εντοπισμού της πηγής σε όρους επιφάνειας εγκεφαλικής περιοχής) εις βάρος της υψηλής υπολογιστικής πολυπλοκότητας. Ο υπολογισμός του πεδίου μολύβδου εξαρτάται επίσης από τον αριθμό των καναλιών ηλεκτροδίων. Ο πραγματικός αριθμός των ηλεκτροδίων συμβάλλει σημαντικά στην ακρίβεια του ESI, με τον περιορισμένο αριθμό καναλιών να οδηγεί σε λανθασμένο εντοπισμό των πηγών [10]. Ωστόσο, ακόμη και αν το ESI χαμηλής πυκνότητας (LD-ESI) είναι λιγότερο ακριβές σε σύγκριση με το ESI υψηλής πυκνότητας (HD-ESI), το LD-ESI είναι πολύ πιο προσιτό και θα μπορούσε να παρέγει αρκετές πληροφορίες δεδομένου του κατάλληλου χρόνου καταγραφής. Συγκεκριμένα, σε νοσοκομειακές εγκαταστάσεις όπου δεν είναι διαθέσιμο το ΗΕΓ υψηλής πυκνότητας, οι μέθοδοι που θα μπορούσαν να αξιοποιήσουν τη δύναμη ενός μικρότερου αριθμού ηλεκτροδίων θα μπορούσαν να ενισχύσουν τη χρηστικότητα του ESI. Στο Σχήμα 4.2, εμφανίζονται διαφορετικά παραδείγματα αριθμού ηλεκτροδίων και σημείων πηγής.

Εξίσου σημαντική είναι και η συν-εγγραφή των αισθητήρων ΗΕΓ, δηλαδή η θέση των ηλεκτροδίων, σε σχέση με το μοντέλο κεφαλής και το πόσο συνεπής είναι η ευθυγράμμιση των ηλεκτροδίων με την πραγματική θέση των ηλεκτροδίων κατά τη διάρκεια των καταγραφών ΗΕΓ. Από αυτή την άποψη, μικρές λανθασμένες τοποθετήσεις κατά τη διάρκεια της υπολογιστικής μοντελοποίησης θα μπορούσαν να οδηγήσουν σε ανακριβείς λύσεις [169]. Οι νέες τεχνικές και οι αγωγοί περιλαμβάνουν τρισδιάστατους σαρωτές προκειμένου να πραγματοποιείται σωστά η ψηφιοποίηση των ηλεκτροδίων και να εξαλείφονται τα σφάλματα καταγραφής από την ανάλυση [170], [171]. Ωστόσο, τα συστήματα αυτά είναι συχνά πολύ ακριβά, ενώ η ευθυγράμμιση των ηλεκτροδίων γίνεται διαδραστικά [162]. Πρέπει επίσης να ληφθεί υπόψη ότι οι σωστοί υπολογισμοί του ESI βασίζονται σε τιμές αγωγιμότητας που παρέγονται στο μοντέλο κεφαλής [169]. Υπό αυτή την προϋπόθεση, η ανατομική μαγνητική τομογραφία θα μπορούσε να συμβάλει στη σωστή επίλυση του εμπρόσθιου (και κατά συνέπεια του αντίστροφου) προβλήματος, καθώς οι πληροφορίες σχετικά με το πάχος και τις αγωγιμότητες κάθε ιστού της κεφαλής είναι απαραίτητες για τον υπολογισμό του εντοπισμού της πηγής [10]. Για το λόγο αυτό, είναι βέλτιστη η χρήση μαγνητικής τομογραφίας για κάθε υποκείμενο προκειμένου να ανακατασκευαστεί το ατομικό μοντέλο κεφαλής, ενσωματώνοντας επιπλέον τις ανισοτροπικές ιδιότητες του κρανίου και της λευκής ουσίας [110], [172].



resolution source model normal resolution source model

c) High-Density EEG with fine resolution source model

Σχήμα 4.2 Παραδείγματα του τρόπου με τον οποίο ο αριθμός των ηλεκτροδίων και των διπόλων (σημεία πηγής) επηρεάζουν την έζοδο ESI. Σε κάθε πίνακα ο υπολογισμένος πίνακας Leadfield (θερμικός χάρτης) εζαρτάται από τα κανάλια ηλεκτροδίων και τα σημεία πηγής

Ωστόσο, σε αρκετές περιπτώσεις, η ατομική μαγνητική τομογραφία δεν είναι διαθέσιμη ή το μοντέλο κεφαλής δεν μπορεί να δημιουργηθεί, λόγω εγκεφαλικών ανωμαλιών και βλαβών, όπως εγκεφαλική δυσπλασία ή όγκοι. Σε τέτοιες συνθήκες, θα μπορούσαν να

χρησιμοποιηθούν προϋπάρχουσες λύσεις προτύπων, όπως τυποποιημένα μοντέλα κεφαλής, αν και τα αποτελέσματα μπορεί να είναι αμφιλεγόμενα ή να ορίζουν ανύπαρκτες περιοχές [173].

Όσον αφορά τις τρέχουσες σύγχρονες μεθόδους που παρουσιάζονται στην παρούσα εργασία, εξακολουθούν να υπάρχουν αρκετοί περιορισμοί. Σχετικά με τις προσεγγίσεις αναγνώρισης προτύπων, το κύριο εμπόδιο για τον αποτελεσματικό εντοπισμό είναι η έλλειψη επισημασμένων πραγματικών δεδομένων που θα χρησιμοποιηθούν για την εκπαίδευση των αρχιτεκτονικών βαθιάς μάθησης. Η επισημείωση πραγματικών δεδομένων ΕΕG με τις πραγματικές θέσεις της πηγής, για παράδειγμα μέσω ταυτόχρονων καταγραφών ΕΕG-fMRI, θα μπορούσε να είναι ένα επαχθές έργο, αλλά ζωτικής σημασίας για την πρόοδο των προσεγγίσεων Βαθιάς Μάθησης στην ΕSI. Αντίθετα, τα μοντέλα Bayesian παρέχουν αυξημένη ακρίβεια εντοπισμού, αν και αυτές οι μέθοδοι είναι εξαιρετικά χρονοβόρες για την επίλυση του αντίστροφου προβλήματος, το οποίο αποτελεί δυσβάσταχτο εμπόδιο για εφαρμογές πραγματικού χρόνου.

Οι πρόσφατες εξελίξεις εστιάζουν προς αυτή την κατεύθυνση, αυξάνοντας την αποδοτικότητα των αλγορίθμων που χρησιμοποιούνται. Οι μελλοντικές πολυτροπικές προσεγγίσεις θα μπορούσαν να συνδυάσουν τα πλεονεκτήματα τόσο των μεθόδων που βασίζονται σε δεδομένα όσο και των μεθόδων που βασίζονται σε μοντέλα, αξιοποιώντας την ακρίβεια εντοπισμού των στατιστικών μεθόδων που εφαρμόζονται στα μοντέλα Bayes με την υπολογιστική ταχύτητα των διαδικασιών Deep Learning.

#### 5. Συμπεράσματα

Στην παρούσα εργασία, παρουσιάζουμε μια ανασκόπηση των πρόσφατων προσεγγίσεων για την επίλυση τόσο του εμπρόσθιου όσο και του αντίστροφου προβλήματος της απεικόνισης πηγών EEG. Επιπλέον, προσπαθήσαμε να συμπυκνώσουμε την πρόοδο των υλοποιήσεων της ESI τόσο σε κλινικές όσο και σε εφαρμογές γνωστικής έρευνας, επισημαίνοντας παράλληλα τις σημερινές προκλήσεις και περιορισμούς. Ως εκ τούτου, αν και οι σύγχρονες εξελίξεις καταδεικνύουν πρόοδο σε μεγάλο βαθμό τόσο στις πτυχές της υπολογιστικής ταχύτητας όσο και της ακρίβειας των διαδικασιών ESI, και περαιτέρω εξελίξεις πρέπει να ενσωματωθούν στα πλαίσια εντοπισμού πηγής, έτσι ώστε οι νέες τεχνολογίες να μπορούν να προσαρμοστούν στα πλαίσια εντοπισμού πηγής, έτσι ώστε οι νέες τεχνολογίες να μπορούν να προσαρμοστούν στα πλαίσια εντοπισμού πηγής, έτσι ώστε οι νέες τεχνολογίες να μπορούν να προσαρμοστούν στα πλαίσια εντοπισμού πηγής, έτσι ώστε οι νέες τεχνολογίες να μπορούν να προσαρμοστούν στα πόρχοντα πρωτόκολλα. Οι μελλοντικές μαθηματικές προσεγγίσεις, οι υπολογιστικές μέθοδοι και η ενσωμάτωση του εντοπισμού της πηγής με άλλες μεθόδους ιατρικής απεικόνισης, θα μπορούσαν να ενισχύσουν την αξία της ESI, με αποτέλεσμα σημαντική πρόοδο στις νευροεπιστήμες, τις τεχνικές αποκατάστασης και την αντίληψη του εγκεφάλου. Με τον τρόπο αυτό, η υψηλή χρονική ανάλυση ενός χωρικά ακριβούς ESI όχι μόνο θα βελτιώσει σημαντικά τη γνωστική αξιολόγηση και τις κλινικές εφαρμογές, αλλά και θα χρησιμεύσει ως ένα προοδευτικό βήμα για τη διερεύνηση των συνόρων του ανθρώπινου νου.

# 5. Κεφάλαιο 5 – Επεξηγησιμότητα με χρήση τεχνικής εντοπισμού πηγής

## 5.1 Εισαγωγή

Τα συστήματα Διεπαφής Εγκεφάλου-Υπολογιστή (Brain-Computer Interface - BCI) επιτρέπουν την άμεση επικοινωνία μεταξύ του εγκεφάλου και εξωτερικών συσκευών, με υποσχόμενες εφαρμογές στη νευροαποκατάσταση, την υποστηρικτική τεχνολογία και την εκπαίδευση νευροανάδρασης [1]. Ωστόσο, οι παραδοσιακές προσεγγίσεις BCI βασίζονται σε δεδομένα ηλεκτροεγκεφαλογραφίας (ΗΕΓ) σε επίπεδο αισθητήρων, τα οποία έχουν εγγενείς περιορισμούς. Ένα βασικό ζήτημα είναι η χαμηλή χωρική ανάλυση του ΗΕΓ τριχωτού της κεφαλής: σήματα από διακριτές περιοχές του εγκεφάλου συγχέονται λόγω της αγωγιμότητας όγκου (το κρανίο και το τριχωτό της κεφαλής δρουν ως χωρικό βαθυπερατό φίλτρο) [2]. Κατά συνέπεια, οι νευρωνικές ενεργοποιήσεις που καταγράφονται στο τριχωτό της κεφαλής δεν μπορούν να εντοπιστούν με βεβαιότητα σε συγκεκριμένες φλοιώδεις περιοχές. Αυτή η αλληλοεπικάλυψη οδηγεί σε αμφίσημες ερμηνείες, όπου οι χάρτες επεξηγησιμότητας αναδεικνύουν μεμονωμένα κανάλια ΗΕΓ αντί για ουσιαστικές εγκεφαλικές περιοχές. Τέτοιες ερμηνείες σε επίπεδο αισθητήρων δυσκολεύονται να ευθυγραμμιστούν με τη νευροανατομική γνώση, εμποδίζοντας τη μετατροπή των ευρημάτων BCI σε κλινικές γνώσεις. Χωρίς τη δυνατότητα συσχέτισης των ενεργοποιήσεων με συγκεκριμένες εγκεφαλικές περιοχές, είναι δύσκολο να αναπτυχθούν στοχευμένες παρεμβάσεις ή να κατανοηθούν πλήρως οι νευρωνικοί μηχανισμοί που διέπουν την κινητική απεικόνιση και άλλες γνωστικές διεργασίες.

Πρόσφατες εξελίξεις προσφέρουν πιθανές λύσεις σε αυτές τις προκλήσεις. Τεχνικές εντοπισμού πηγής όπως ο διαμορφωτής δέσμης Γραμμικά Περιορισμένης Ελάχιστης Διασποράς (Linearly Constrained Minimum Variance - LCMV) προβάλλουν τα σήματα ΗΕΓ στον όγκο του εγκεφάλου για να εκτιμήσουν την προέλευση της νευρωνικής δραστηριότητας [3]. Ο LCMV παράγει χωρικά φίλτρα που διέρχονται δραστηριότητα από μια επιθυμητή θέση ενώ εξασθενούν άλλες, βελτιώνοντας έτσι τη χωρική εξειδίκευση στην ανακατασκευή πηγής ΗΕΓ [3]. Ανακατασκευάζοντας το ΗΕΓ στον φλοιϊκό χώρο πηγής, λαμβάνουμε μια αναπαράσταση της νευρωνικής δραστηριότητας με πολύ υψηλότερη ανατομική πιστότητα. Επιπλέον, ο συνδυασμός του εντοπισμού πηγής με σχήματα κατάτμησης εγκεφαλικών περιοχών μπορεί να αποφέρει πιο ερμηνεύσιμα χαρακτηριστικά. Για παράδειγμα, ο άτλαντας Desikan–Killiany χωρίζει τον φλοιό σε 68 ανατομικά καθορισμένες περιοχές ενδιαφέροντος (Regions of Interest - ROIs) [4]. Τέτοιοι άτλαντες χρησιμοποιούνται ευρέως στη νευροαπεικόνιση λόγω της ανατομικής τους εγκυρότητας, επιτρέποντάς μας να συγκεντρώνουμε σήματα ΗΕΓ σε επίπεδο πηγής σε χρονοσειρές συγκεκριμένων περιοχών. Αυτή η προσέγγιση μπορεί να γεφυρώσει το γάσμα μεταξύ μοντέλων που καθοδηγούνται από δεδομένα και της καθιερωμένης φλοιϊκής ανατομίας.

Ο στόχος αυτής της μελέτης είναι να αξιοποιήσει τμηματοποιημένα δεδομένα ΗΕΓ εντοπισμένης πηγής για να ενισχύσει την ερμηνευσιμότητα και την κλινική συνάφεια της επεξηγησιμότητας των νευρωνικών δικτύων στην ταξινόμηση κινητικής απεικόνισης. Προτείνουμε ένα νέο πλαίσιο που επεξεργάζεται ακατέργαστα δεδομένα ΗΕΓ μέσω

εντοπισμού πηγής και στη συνέχεια τμηματοποιεί τα δεδομένα σε 68 φλοιϊκές ROIs. Αυτή η μέθοδος στοχεύει: (1) Να ενισχύσει τη χωρική εξειδίκευση εντοπίζοντας με ακρίβεια τις νευρωνικές ενεργοποιήσεις εντός διακριτών φλοιϊκών περιοχών, (2) Να βελτιώσει την ερμηνευσιμότητα των εξηγήσεων του μοντέλου χαρτογραφώντας τις αποδόσεις του νευρωνικού δικτύου σε γνωστές εγκεφαλικές περιοχές, και (3) Να ευθυγραμμίσει τα ευρήματα BCI με τη νευροεπιστημονική βιβλιογραφία, διευκολύνοντας την ευκολότερη επικύρωση και εφαρμογή των αποτελεσμάτων. Προηγούμενες μελέτες έχουν εφαρμόσει μεθόδους επεξηγήσιμης ΤΝ σε δεδομένα ΗΕΓ σε επίπεδο αισθητήρων σε BCI, για παράδειγμα χρησιμοποιώντας χάρτες έμφασης (saliency maps) ή χάρτες συνάφειας (relevance maps) για την ερμηνεία ταξινομητών βαθιάς μάθησης [5]. Ενώ αυτές οι προσπάθειες καταδεικνύουν την αξία της ερμηνευσιμότητας στα BCI, περιορίζονται θεμελιωδώς από την έλλειψη χωρικής ακρίβειας σε επίπεδο αισθητήρων. Οι προκύπτουσες εξηγήσεις μπορεί να είναι δύσκολο να εμπιστευτεί κανείς ή να τις συσχετίσει με την υποκείμενη νευροφυσιολογία [5]. Εξ όσων γνωρίζουμε, καμία προηγούμενη εργασία δεν έχει διερευνήσει τη χρήση ανατομικά θεμελιωμένων εισόδων ΗΕΓ σε επίπεδο πηγής για την εξήγηση μοντέλων BCI. Καλύπτοντας αυτό το κενό, δείχνουμε ότι η προσέγγισή μας αποφέρει βαθύτερες γνώσεις σχετικά με τους νευρωνικούς συσχετισμούς της κινητικής απεικόνισης και βελτιώνει την κλινική ερμηνευσιμότητα των συστημάτων BCI.

Συνοπτικά, ενσωματώνουμε προηγμένες τεχνικές εντοπισμού πηγής ΗΕΓ με τεχνικές επεξηγησιμότητας νευρωνικών δικτύων – συγκεκριμένα Χάρτες Έμφασης (Saliency Maps), Ολοκληρωμένες Κλίσεις (Integrated Gradients) και τιμές SHAP – για να δημιουργήσουμε μια πιο ανατομικά ουσιαστική ερμηνεία της αποκωδικοποίησης της κινητικής απεικόνισης. Αυτό το πλαίσιο ενισχύει την ερμηνευσιμότητα των BCI και βοηθά στη γεφύρωση του χάσματος μεταξύ υπολογιστικών μοντέλων και νευροανατομικής κατανόησης, ένα βήμα προς πιο διαφανή και κλινικά συναφή συστήματα BCI.

## 5.2 Σχετικές Εργασίες

Η έρευνα στα BCI έχει σημειώσει σημαντική πρόοδο στην ταξινόμηση κινητικής απεικόνισης βάσει ΗΕΓ τα τελευταία πέντε χρόνια. Εδώ εξετάζουμε πρόσφατες μελέτες που, όπως και η δική μας εργασία, χρησιμοποιούν το σύνολο δεδομένων BCI Competition IV-2a για την αξιολόγηση προηγμένων μεθόδων ταξινόμησης και ερμηνείας. Οι παραδοσιακές προσεγγίσεις μηχανικής μάθησης (π.χ., Κοινά Χωρικά Πρότυπα με γραμμικούς ταξινομητές) έχουν όλο και περισσότερο ξεπεραστεί από μοντέλα βαθιάς μάθησης. Πρώιμες αρχιτεκτονικές βαθιών CNN όπως το Deep ConvNet [6] και το ελαφρύ EEGNet [7] έχουν γίνει καθιερωμένα σημεία αναφοράς στον τομέα. Βασιζόμενοι σε αυτά, οι ερευνητές έχουν αναπτύξει νέες αρχιτεκτονικές για τη βελτίωση της ακρίβειας και της ευρωστίας στο BCI IV-2a:

• **EEG-TCNet** (2020): Οι Ingolfsson et al. εισήγαγαν το EEG-TCNet, ένα Χρονικό Συνελικτικό Δίκτυο (Temporal Convolutional Network) βελτιστοποιημένο για σήματα ΗΕΓ. Το EEG-TCNet πέτυχε περίπου 77,4% ακρίβεια 4 κλάσεων στο BCI IV-2a με πολύ λιγότερες παραμέτρους από προηγούμενα CNNs [8], καθιστώντας το κατάλληλο για ενσωματωμένα συστήματα BCI πραγματικού χρόνου. Η επιτυχία του κατέδειξε την αποτελεσματικότητα της

εστίασης στη χρονική δυναμική με διεσταλμένες συνελίξεις (dilated convolutions) για την ταξινόμηση ΚΑ (Κινητικής Απεικόνισης).

• EEG-Inception (2021): Οι Santamaria-Vazquez et al. πρότειναν το EEG-Inception, ένα βαθύ δίκτυο βασισμένο στην ιδέα της ενότητας Inception, για την καταγραφή χαρακτηριστικών πολλαπλών κλιμάκων σε χρονοσειρές ΗΕΓ. Το EEG-Inception ξεπέρασε αρκετές προηγούμενες προσεγγίσεις (συμπεριλαμβανομένων παραλλαγών του CSP και υβριδίων CNN-LSTM) στο BCI IV-2a, τονίζοντας ότι οι συνελίξεις πολλαπλών κλιμάκων μπορούν να εξάγουν πιο διακριτικά χαρακτηριστικών χρονικών κλιμάκων άσκησε επιρροή σε μεταγενέστερες αρχιτεκτονικές.

• Προσεγγίσεις Μεταφοράς Μάθησης (Transfer Learning): Για την αντιμετώπιση της μεταβλητότητας μεταξύ συνεδριών ή υποκειμένων, οι ερευνητές έχουν διερευνήσει τη μεταφορά μάθησης στο BCI IV-2a. Για παράδειγμα, οι Solorzano et al. (2021) χρησιμοποίησαν έναν αυτοκωδικοποιητή για προ-εκπαίδευση σε δεδομένα πολλαπλών υποκειμένων και στη συνέχεια τελειοποίησαν έναν ταξινομητή για κάθε υποκείμενο, βελτιώνοντας την ακρίβεια ταξινόμησης ΚΑ μεταξύ υποκειμένων σε σύγκριση με την εκπαίδευση από την αρχή. Τέτοιες προσεγγίσεις αξιοποιούν την κοινή δομή στα χαρακτηριστικά ΗΕΓ για τη μείωση του χρόνου βαθμονόμησης [10]. Ομοίως, πλαίσια προσαρμογής πεδίου (domain adaptation) (π.χ., ευθυγράμμιση Riemannian ή προσαρμοστική κανονικοποίηση δέσμης - adaptive batch normalization) έχουν εφαρμοστεί στο BCI IV-2a για τον μετριασμό των διαφορών από συνεδρία σε συνεδρία.

• CNNs Βασισμένα στην Προσοχή (Attention-Based CNNs): Η ενσωμάτωση μηχανισμών προσοχής έχει αποδειχθεί αποτελεσματική σε πρόσφατα μοντέλα. Οι Altaheri et al. (2022) ανέπτυξαν το ATCNet, το οποίο ενσωματώνει αυτο-προσοχή πολλαπλών κεφαλών (multihead self-attention) με έναν χρονικό συνελικτικό κορμό [11]. Το ATCNet πέτυχε απόδοση αιχμής (state-of-the-art) στο BCI IV-2a, καταδεικνύοντας ότι τα επίπεδα προσοχής μπορούν να βοηθήσουν το δίκτυο να εστιάσει σε σημαντικά χρονοσυχνικά στοιχεία του ΗΕΓ. Οι Hu et al. (2023) επέκτειναν αυτή την ιδέα με το MSATNet, ένα CNN διπλού κλάδου συνδυασμένο με μια ενότητα Transformer για την καταγραφή καθολικών χρονικών εξαρτήσεων [12]. Αυτά τα μοντέλα βασισμένα στην προσοχή βελτίωσαν σημαντικά την ακρίβεια ταξινόμησης, υποδεικνύοντας το όφελος της προσαρμοστικής στάθμισης σημαντικών χαρακτηριστικών ΗΕΓ.

• Υβριδικά και Συνδυαστικά Movτέλα (Hybrid and Ensemble Models): Αρκετές εργασίες έχουν συνδυάσει διαφορετικές ενότητες βαθιάς μάθησης για την ενίσχυση της απόδοσης στο σύνολο δεδομένων 2a. Για παράδειγμα, οι Hui et al. (2025) συνδύασαν CNNs με χρονικά επίπεδα LSTM και προσοχή για την καταγραφή τόσο χωρικών όσο και χρονικών χαρακτηριστικών του ΗΕΓ ΚΑ [13], ενώ οι Xie et al. (2024) πρότειναν ένα δίκτυο τριών κλάδων (BFATCNet) που συνδυάζει συνελικτικά μπλοκ και μπλοκ προσοχής για ολοκληρωμένη εξαγωγή χαρακτηριστικών [14]. Οι Nguyen et al. (2024) συνδύασαν ομοίως παράλληλες ενότητες προσοχής με κορμούς EEGNet και TCN σε μια αρχιτεκτονική συνόλου (ensemble) [15]. Αυτές οι πολύπλοκες αρχιτεκτονικές αναφέρουν ανταγωνιστικά αποτελέσματα, αν και με κόστος υψηλότερης πολυπλοκότητας μοντέλου.

Δείκτες Απόδοσης (Performance Benchmarks): Η ακρίβεια ταξινόμησης στο BCI IV-2a έχει αυξηθεί σταθερά με αυτές τις καινοτομίες. Ως αναφορά, η αρχική γραμμή βάσης CSP+LDA ήταν περίπου ~65% για KA 4 κλάσεων. Πρόσφατα βαθιά δίκτυα υπερβαίνουν πλέον συστηματικά το 75–80%. Αξιοσημείωτα, ένα CNN διπλού κλάδου με προσοχή (CIACNet) από τους Liao et al. (2025) πέτυχε ακρίβεια 85,15% στο BCI IV-2a [16] – μία από τις υψηλότερες που έχουν αναφερθεί σε αυτό το σύνολο δεδομένων. Αυτό το μοντέλο χρησιμοποίησε μια βελτιωμένη ενότητα προσοχής συνελικτικού μπλοκ (Convolutional Block Attention Module - CBAM) και σύντηξη χαρακτηριστικών πολλαπλών επιπέδων, υπογραμμίζοντας την αξία της προσοχής και της μάθησης πολλαπλών κλιμάκων. Τέτοια κέρδη απόδοσης καταδεικνύουν πόσο έχουν προχωρήσει οι σύγχρονες τεχνικές στην αξιοπιστία των BCI.

Σημαντικό είναι ότι, παράλληλα με την ακρίβεια, πρόσφατες μελέτες δίνουν έμφαση στην ερμηνευσιμότητα. Για παράδειγμα, οι Zhou et al. (2022) εφάρμοσαν Grad-CAM και SHAP για να προσδιορίσουν τις εγκεφαλικές περιοχές και τις ζώνες συχνοτήτων που ήταν πιο σχετικές για τις αποφάσεις του CNN τους, κάνοντας παραλληλισμούς με νευροφυσιολογικά πρότυπα. Ομοίως, η επεξηγήσιμη βαθιά μάθηση για BCI είναι ένας αναπτυσσόμενος τομέας – η εργασία μας συμβάλλει σε αυτό συνδυάζοντας τον εντοπισμό πηγής με την επεξηγησιμότητα, ενώ οι περισσότερες προηγούμενες εργασίες ερμηνεύουν μοντέλα στον χώρο των αισθητήρων, παρέχουμε νέες προοπτικές για την κατανόηση και την επικύρωση των μοντέλων BCI.

Συνολικά, η βιβλιογραφία δείχνει μια σαφή τάση προς εξελιγμένα μοντέλα βαθιάς μάθησης (CNNs, RNNs, Transformers και υβριδικά) εκπαιδευμένα και αξιολογημένα σε κοινά σημεία αναφοράς όπως το BCI Competition IV-2a. Αυτά τα μοντέλα, συμπεριλαμβανομένων όσων εξετάστηκαν παραπάνω, αποτελούν την τελευταία λέξη της τεχνολογίας (state of the art) στην οποία βασίζεται η προσέγγισή μας. Η συμβολή μας είναι μοναδική καθώς ενισχύουμε έναν από αυτούς τους αγωγούς (pipelines) βαθιάς μάθησης υψηλής απόδοσης με ανάλυση σε επίπεδο πηγής και ανατομικά θεμελιωμένη ερμηνευσιμότητα, με στόχο να διευρύνουμε τα όρια όχι μόνο στην απόδοση ταξινόμησης αλλά και στη διαφάνεια και τη νευροεπιστημονική συνάφεια των μοντέλων BCI.

## 5.3 Μεθοδολογία

#### Σύνολο Δεδομένων και Προεπεξεργασία

Αξιολογούμε την προσέγγισή μας στο σύνολο δεδομένων BCI Competition IV 2a [17], ένα καθιερωμένο σημείο αναφοράς για BCI κινητικής απεικόνισης. Αυτό το σύνολο δεδομένων περιέχει καταγραφές ΗΕΓ από 9 υποκείμενα, καθένα από τα οποία εκτελεί τέσσερις εργασίες κινητικής απεικόνισης (KA): αριστερό χέρι, δεξί χέρι, και τα δύο πόδια, και κινήσεις γλώσσας. Το ΗΕΓ καταγράφηκε από 22 ηλεκτρόδια Ag/AgCl τοποθετημένα σύμφωνα με το διεθνές σύστημα 10–20 (τυπικές θέσεις), με ρυθμό δειγματοληψίας 250 Hz. Επιπλέον, καταγράφηκαν 3 κανάλια ηλεκτροοφθαλμογραφήματος (electrooculogram - EOG) για την παρακολούθηση τεχνουργημάτων [17].

Εφαρμόσαμε μια σειρά βημάτων προεπεξεργασίας για την ενίσχυση της ποιότητας του σήματος. Πρώτον, πραγματοποιήσαμε αφαίρεση τεχνουργημάτων χρησιμοποιώντας Ανάλυση Ανεξάρτητων Συνιστωσών (Independent Component Analysis - ICA). Η ICA είναι μια ευρέως χρησιμοποιούμενη τεχνική για τον διαχωρισμό πηγών ΗΕΓ και την αφαίρεση συνιστωσών θορύβου (όπως βλεφαρισμοί και μυϊκή δραστηριότητα) χωρίς απόρριψη χρήσιμων δεδομένων [18], [19]. Οι συνιστώσες που αντιπροσώπευαν οφθαλμικά ή μυϊκά τεχνουργήματα αναγνωρίστηκαν από τις χαρακτηριστικές τους τοπογραφίες και χρονοσειρές και στη συνέχεια αφαιρέθηκαν από το ΗΕΓ, παρόμοια με την προσέγγιση των Fatourechi et al. (2007) [20]. Στη συνέχεια, τα σήματα ΗΕΓ φιλτραρίστηκαν ζωνοπερατά στην περιοχή 8–30 Ηz χρησιμοποιώντας ένα φίλτρο πεπερασμένης κρουστικής απόκρισης (finite impulse response filter). Αυτή η περιοχή συλλαμβάνει τους ρυθμούς μι (~8–13 Hz) και βήτα (~13–30 Hz), οι οποίοι είναι γνωστό ότι παρουσιάζουν αποσυγχρονισμό σχετιζόμενο με γεγονότα (event-related desynchronization) κατά τη διάρκεια της κινητικής απεικόνισης [16]. Η εστίαση σε αυτές τις ζώνες τονίζει τα ταλαντωτικά χαρακτηριστικά που είναι πιο συναφή με τις εργασίες ΚΑ.

[Εικόνα Ι. Χρονικό διάγραμμα των δοκιμών.]

Το συνεχές ΗΕΓ στη συνέχεια τμηματοποιήθηκε σε εποχές ευθυγραμμισμένες με τις δοκιμές κινητικής απεικόνισης. Κάθε εποχή δοκιμής εκτείνεται από 0.5 δευτερόλεπτα μετά την έναρξη του ερεθίσματος (cue onset) έως 2 δευτερόλεπτα μετά την έναρξη του ερεθίσματος, αποδίδοντας ένα παράθυρο 2 δευτερολέπτων ανά προσπάθεια απεικόνισης. Αυτό το παράθυρο εποχής συλλαμβάνει τόσο την προπαρασκευαστική (πριν το ερέθισμα) όσο και την εκτελεστική φάση της κινητικής απεικόνισης. Για τη μείωση της μεταβλητότητας της γραμμής βάσης, πραγματοποιήσαμε διόρθωση γραμμής βάσης αφαιρώντας το μέσο σήμα του διαστήματος 0.5 δευτερολέπτων πριν το ερέθισμα από ολόκληρη την εποχή. Το αποτέλεσμα είναι ότι κάθε εποχή κανονικοποιείται ως προς τη γραμμή βάσης της πριν την απεικόνιση, βελτιώνοντας τη συνέπεια μεταξύ των δοκιμών.

#### Εντοπισμός Πηγής με Διαμορφωτές Δέσμης LCMV

Για να ξεπεράσουμε την περιορισμένη χωρική ανάλυση του ΗΕΓ τριχωτού της κεφαλής, χρησιμοποιήσαμε εντοπισμό πηγής χρησιμοποιώντας έναν διαμορφωτή δέσμης LCMV. Ο διαμορφωτής δέσμης LCMV είναι μια μέθοδος χωρικού φιλτραρίσματος που υπολογίζει ένα σύνολο βαρών για τους αισθητήρες ΗΕΓ ώστε να εκτιμήσει τη δραστηριότητα που προέρχεται από μια δεδομένη θέση στον εγκέφαλο, καταστέλλοντας ταυτόχρονα τις παρεμβολές από άλλες θέσεις [3], [21]. Στην πράξη, κατασκευάζουμε ένα ρεαλιστικό μοντέλο κεφαλής και το χρησιμοποιούμε για να εξαγάγουμε το εμπρόσθιο μοντέλο (πίνακας πεδίου αγωγής - leadfield matrix) που συσχετίζει τη μοναδιαία δραστηριότητα πηγής σε διάφορες θέσεις του εγκεφάλου με τις μετρήσεις σε κάθε ηλεκτρόδιο ΗΕΓ. Για το μοντέλο κεφαλής, χρησιμοποιήσαμε το πρότυπο εγκεφάλου Montreal Neurological Institute (MNI) 152 και ένα μοντέλο αγωγιμότητας όγκου τριών κελυφών (για το τριχωτό της κεφαλής, το κρανίο και τον εγκέφαλο) βασισμένο στη Μέθοδο Οριακών Στοιχείων (Boundary Element Method). Αυτό το μοντέλο προσεγγίζει την αγώγιμη γεωμετρία της κεφαλής και χρησιμοποιείται συνήθως όταν δεν είναι διαθέσιμη μαγνητική τομογραφία (MRI) ειδικά για το υποκείμενο. Χρησιμοποιώντας το μοντέλο κεφαλής και τις συντεταγμένες των ηλεκτροδίων, υπολογίσαμε τον πίνακα πεδίου αγωγής για ένα πυκνό πλέγμα υποψήφιων θέσεων πηγής (δίπολα) στον φλοιό. Το πεδίο αγωγής καθορίζει πώς κάθε μοναδιαία πηγή συμβάλλει σε κάθε κανάλι ΗΕΓ. Δεδομένου του πεδίου αγωγής και της συνδιακύμανσης των αισθητήρων των προεπεξεργασμένων δεδομένων μας, υπολογίσαμε τα βάρη του διαμορφωτή δέσμης LCMV για κάθε σημείο του πλέγματος. Διαισθητικά, το διάνυσμα βάρους του διαμορφωτή δέσμης για μια δεδομένη θέση επιλέγεται έτσι ώστε να διέρχεται σήματα που προέρχονται από αυτή τη θέση με μοναδιαίο κέρδος, ελαχιστοποιώντας παράλληλα την ισχύ εξόδου (διασπορά) που συνεισφέρουν όλες οι άλλες πηγές. Αυτή η βελτιστοποίηση γραμμικού περιορισμού αποδίδει ένα σύνολο χωρικών φίλτρων, ένα ανά θέση. Εφαρμόσαμε αυτά τα φίλτρα στις εποχές ΗΕΓ για να λάβουμε χρονοσειριακές εκτιμήσεις της νευρωνικής δραστηριότητας σε κάθε υποθετική θέση πηγής (συχνά αποκαλούμενες «εικονικοί αισθητήρες»). Στην πραγματικότητα, αυτό μετασχηματίζει τα δεδομένα από τον χώρο των αισθητήρων στον χώρο της πηγής, αποδίδοντας μια ογκομετρική χρονοσειριακή αναπαράσταση της εγκεφαλικής δραστηριότητας για κάθε δοκιμή [21].

#### Στρατηγική Κατάτμησης (Parcellation)

Μετά τον εντοπισμό πηγής, αντιμετωπίσαμε την πρόκληση της ερμηνείας μιας πολύ υψηλής διάστασης εξόδου σε επίπεδο πηγής (με χιλιάδες εικονικούς αισθητήρες). Για να την αποστάξουμε σε μια πιο διαχειρίσιμη και ουσιαστική αναπαράσταση, εφαρμόσαμε μια στρατηγική φλοιϊκής κατάτμησης. Χρησιμοποιήσαμε τον άτλαντα Desikan–Killiany, ο οποίος χωρίζει τον εγκεφαλικό φλοιό σε 68 ανατομικά καθορισμένες περιοχές ενδιαφέροντος (34 ανά ημισφαίριο) [4]. Αυτός ο άτλαντας παρέχει ετικέτες για κύριες έλικες και φλοιϊκές περιοχές (π.χ., προκινητικός φλοιός, σωματοαισθητικές περιοχές, κ.λπ.) και χρησιμοποιείται ευρέως ως ανατομικό πλαίσιο στη νευροεπιστήμη.

Για κάθε υποκείμενο, τα δεδομένα ΗΕΓ εντοπισμένης πηγής χαρτογραφήθηκαν στις περιοχές Desikan-Killiany. Συγκεκριμένα, κάθε εικονικός αισθητήρας του διαμορφωτή δέσμης (σημείο πλέγματος) αντιστοιχίστηκε στην αντίστοιχη ROI του βάσει της ανατομικής του θέσης. Στη συνέχεια, για κάθε ROI, συγκεντρώσαμε τις χρονοσειρές πηγής όλων των σημείων πλέγματος εντός αυτής της περιοχής υπολογίζοντας τον μέσο όρο τους. Αυτό παρήγαγε μια ενιαία αντιπροσωπευτική χρονοσειρά για καθεμία από τις 68 φλοιϊκές περιοχές ανά δοκιμή. Το αποτέλεσμα αυτής της κατάτμησης είναι ότι κάθε δοκιμή ΗΕΓ αναπαρίσταται πλέον από έναν πίνακα 68×T (68 περιοχές επί Τ χρονικά σημεία, με T=500 δείγματα για 2 δευτερόλεπτα στα 250 Hz). Αυτή η προσέγγιση προσφέρει πολλά πλεονεκτήματα. Πρώτον, παρέχει ανατομική θεμελίωση – κάθε χαρακτηριστικό αντιστοιχεί πλέον σε μια συγκεκριμένη εγκεφαλική περιοχή με γνωστή λειτουργική σημασία, διευκολύνοντας την ερμηνεία. Δεύτερον, επιτυγχάνει δραστική μείωση διαστάσεων (από γιλιάδες χαρακτηριστικά σε επίπεδο voxel σε 68 χαρακτηριστικά σε επίπεδο περιοχής), η οποία βελτιώνει τον λόγο σήματος προς θόρυβο και μειώνει την πολυπλοκότητα του μοντέλου. Τρίτον, επιτρέπει την άμεση σύγκριση των αποτελεσμάτων μας με την υπάρχουσα βιβλιογραφία (π.χ. προσδιορισμός ποιων φλοιϊκών περιοχών εμπλέκονται σε μια εργασία, αντί για ποια κανάλια ΗΕΓ). Κατατμίζοντας τη δραστηριότητα της πηγής, διασφαλίζουμε ότι οι επακόλουθες αναλύσεις νευρωνικών δικτύων και οι εξηγήσεις τους μπορούν να συσγετιστούν άμεσα με ουσιαστικές εγκεφαλικές περιογές αντί για θέσεις αισθητήρων.

[Εικόνα ΙΙ. Η διαδικασία εντοπισμού πηγής και κατάτμησης των σημάτων ΗΕΓ]

#### Αρχιτεκτονική Νευρωνικού Δικτύου: CNN1D\_Custom

Για την ταξινόμηση των εργασιών κινητικής απεικόνισης από το τμηματοποιημένο ΗΕΓ σε επίπεδο πηγής, σχεδιάσαμε ένα προσαρμοσμένο μονοδιάστατο συνελικτικό νευρωνικό δίκτυο (CNN) που αναφέρεται ως CNN1D\_Custom. Τα συνελικτικά νευρωνικά δίκτυα έχουν επιδείξει απόδοση αιχμής (state-of-the-art) στην ταξινόμηση BCI βάσει ΗΕΓ, ξεπερνώντας πολλές παραδοσιακές προσεγγίσεις [5]. Αξιοσημείωτα, τα βαθιά CNNs μπορούν να μάθουν αυτόματα χωροχρονικά χαρακτηριστικά από ακατέργαστα σήματα ΗΕΓ, όπως έδειξαν οι Schirrmeister et al. (2017) [6] και το συμπαγές μοντέλο EEGNet των Lawhern et al. (2018) [7]. Παρακινούμενοι από αυτές τις επιτυχίες, το CNN1D\_Custom μας είναι προσαρμοσμένο για να εκμεταλλεύεται τη χωροχρονική δομή της εισόδου φλοιϊκών χρονοσειρών 68 καναλιών.

**Αναπαράσταση εισόδου:** Κάθε δείγμα εισόδου στο δίκτυο είναι ένας πίνακας 68×500 (68 περιοχές × 2 δευτερόλεπτα δεδομένων). Το αντιμετωπίζουμε ως 68 μονομεταβλητές χρονοσειρές. Για την εκπαίδευση του δικτύου, κάθε είσοδος φέρει ετικέτα με μία από τις τέσσερις κλάσεις (ΚΑ αριστερού χεριού, δεξιού χεριού, ποδιών, γλώσσας).

Αρχιτεκτονική CNN: Το δίκτυο ξεκινά με δύο διαδοχικά συνελικτικά επίπεδα που λειτουργούν κατά μήκος της χρονικής διάστασης για κάθε περιοχή. Το πρώτο συνελικτικό επίπεδο χρησιμοποιεί 32 φίλτρα με μέγεθος πυρήνα 3 (χρονικό παράθυρο 3 δειγμάτων), που εφαρμόζονται με ίδια επένδυση (same padding). Αυτό το επίπεδο μπορεί να ανιχνεύσει τοπικά χρονικά πρότυπα στο σήμα κάθε περιοχής (για παράδειγμα, εκρήξεις ρυθμικής δραστηριότητας ή γεγονότα ERD/ERS). Το δεύτερο συνελικτικό επίπεδο χρησιμοποιεί 64 φίλτρα (μέγεθος πυρήνα 3) για να συλλάβει πιο πολύπλοκα πρότυπα και αλληλεπιδράσεις μεταξύ περιοχών. Χρησιμοποιούμε ενεργοποίηση Διορθωμένης Γραμμικής Μονάδας (Rectified Linear Unit - ReLU) μετά από κάθε συνέλιξη για να εισαγάγουμε μη γραμμικότητα. Μετά τις συνελίξεις, ένα επίπεδο μέγιστης συγκέντρωσης (max-pooling layer) (μέγεθος συγκέντρωσης 2) κατά μήκος του χρόνου χρησιμοποιείται για την υποδειγματοληψία των χαρτών χαρακτηριστικών, μειώνοντας τη χρονική ανάλυση διατηρώντας παράλληλα τις πιο προεξέχουσες ενεργοποιήσεις. Αυτό παρέχει επίσης κάποια μετατοπιστική αναλλοίωτη στη χρονική στιγμή των χαρακτηριστικών και βοηθά στην πρόληψη της υπερπροσαρμογής (overfitting). Μετά τη συνέλιξη και τη συγκέντρωση, οι χάρτες χαρακτηριστικών από όλες τις 68 περιοχές συνενώνονται (concatenated) και ισοπεδώνονται (flattened). Αυτό το ισοπεδωμένο διάνυσμα τροφοδοτείται σε πλήρως συνδεδεμένα πυκνά επίπεδα (fully connected dense layers) για την ενσωμάτωση πληροφοριών από όλες τις περιοχές. Χρησιμοποιούμε δύο πυκνά επίπεδα μεγέθους 128 και 64, καθένα ακολουθούμενο από ReLU. Αυτά τα επίπεδα μαθαίνουν συνδυασμούς χαρακτηριστικών από διαφορετικές περιοχές και χρονικές περιόδους, μαθαίνοντας αποτελεσματικά διακριτικά πρότυπα υψηλού επιπέδου που σχετίζονται με κάθε κλάση κινητικής απεικόνισης. Τέλος, το δίκτυο εξάγει πιθανότητες κλάσεων μέσω ενός επιπέδου softmax με 4 μονάδες. Η κλάση με την υψηλότερη πιθανότητα λαμβάνεται ως η προβλεπόμενη κλάση ΚΑ.

Διαδικασία εκπαίδευσης: Χρησιμοποιήσαμε την κατηγορική διασταυρούμενη εντροπία (categorical cross-entropy) ως συνάρτηση απώλειας (κατάλληλη για πολυκπασική ταξινόμηση) και τον βελτιστοποιητή Adam για στοχαστική κάθοδο κλίσης (stochastic gradient descent). Ο ρυθμός μάθησης ορίστηκε σε 0.001 και η εκπαίδευση συνεχίστηκε για μέγιστο 100 εποχές. Χρησιμοποιήσαμε πρώιμη διακοπή (early stopping) σε ένα σύνολο επικύρωσης για την πρόληψη της υπερπροσαρμογής – εάν η απώλεια επικύρωσης δεν βελτιωνόταν για 10 συνεχόμενες εποχές, η εκπαίδευση διακοπτόταν. Για την αύξηση του σχετικά μικρού συνόλου δεδομένων, χρησιμοποιήσαμε τεχνικές όπως τυχαία χρονική αποκοπή (random temporal cropping) ή ελαφρά μετατόπιση των δοκιμών, ανάλογα με τις ανάγκες. Η εκπαίδευση και η αξιολόγηση του δικτύου έγιναν με τρόπο ειδικό για κάθε υποκείμενο (εκπαίδευση και δοκιμή στα δεδομένα του ίδιου υποκειμένου χρησιμοποιώντας διασταυρούμενη επικύρωση - cross-validation), δεδομένου ότι τα δεδομένα του διαγωνισμού BCI συνήθως αξιολογούν την απόδοση εντός του υποκειμένου.

#### 5.4 Αποτελέσματα επεξηγησημότητας μοντέλου εντοπισμού πηγής και τμηματοποίησης (εφαρμογή σε Brain Computer Interface)

Όσον αφορά την εφαρμιγη source localization και τμηματοποίησης για την βελτίωση της ερμηνευσιμότητας, παρουσιάζουμε τα αποτελέσματα σε δύο μορφές: Χάρτες θερμότητας και χάρτες εγκεφάλου. Οι χάρτες θερμότητας είναι δισδιάστατα διαγράμματα (περιογές × γρόνος) όπου η ένταση του χρώματος αντικατοπτρίζει τη σημασία του χαρακτηριστικού (από Saliency, IG ή SHAP) όπως φαίνεται και στο Σχήμα 5.2. Αυτό μας επιτρέπει να παρατηρήσουμε γρονικά μοτίβα σημαντικότητας σε κάθε περιογή. Τέτοια μοτίβα μπορούν να συγκριθούν με αναμενόμενα φυσιολογικά φαινόμενα (όπως η χρονική πορεία του αποσυγχρονισμού που σχετίζεται με γεγονότα). Επιπλέον, προβάλλονται οι είσοδοι χαρακτηριστικών σε μια πρότυπη επιφάνεια του εγκεφάλου (fsaverage) για να δημιουργηθούν διαισθητικοί χάρτες του εγκεφάλου. Ένα διαδραστικό διάγραμμα εγκεφάλου χρησιμοποιείται για την εμφάνιση των τιμών SHAP για κάθε χρονική στιγμή, όπως φαίνεται στο Σχήμα 5.3. Αυτοί οι εγκεφαλικοί χάρτες δείχνουν άμεσα ποιες περιοχές του φλοιού είχαν μεγαλύτερη επιρροή στην απόφαση του μοντέλου. Δημιουργήσαμε ξεχωριστούς χάρτες για κάθε κατηγορία και μέθοδο. Για παράδειγμα, ένας εγκεφαλικός γάρτης με βάση το IG για την απεικόνιση του αριστερού χεριού μπορεί να αναδείξει τον δεξιό κινητικό φλοιό (αντίθετη περιογή του γεριού M1) καθώς και τη συμπληρωματική κινητική περιογή, σύμφωνα με τη νευροφυσιολογία. Μια τέτοια συμφωνία μεταξύ της εξήγησης του μοντέλου και των γνωστών κινητικών δικτύων θα αύξανε την εμπιστοσύνη στο μοντέλο- οι αποκλίσεις θα μπορούσαν να υποδηλώνουν τεχνουργήματα ή ψευδείς συντομεύσεις που έμαθε το μοντέλο (π.χ. υπερβολική εξάρτηση από τα σήματα της μετωπικής κίνησης των ματιών).

Η ακρίβεια του μοντέλου ήταν ένα ισχυρό 76%, στα ίδια επίπεδα με τα περισσότερα από τα αντίστοιχα μοντέλα τελευταίας τεχνολογίας. Ωστόσο, η απλότητά του μας επιτρέπει να χρησιμοποιήσουμε και να αξιοποιήσουμε τα μέτρα επεξηγηματικότητας που περιγράφηκαν προηγουμένως. Αναλύσαμε τις εξηγήσεις SHAP για ένα 1D CNN που εκπαιδεύτηκε σε EEG εντοπισμένης πηγής, εστιάζοντας στις τέσσερις κατηγορίες κινητικής απεικόνισης (αριστερό χέρι, δεξί χέρι, πόδια, γλώσσα). Για κάθε κλάση, οι υψηλότερες θετικές τιμές SHAP υποδεικνύουν περιοχές του εγκεφάλου των οποίων η δραστηριότητα υποστηρίζει την εν λόγω κλάση (χαρακτηριστικά υπέρ της κλάσης), ενώ οι αρνητικές τιμές SHAP προσδιορίζουν περιοχές των οποίων η δραστηριότητα αντιτίθεται στην εν λόγω κλάση (χαρακτηριστικά κατά

της κλάσης). Αυτές οι περιοχές που εντοπίστηκαν από το μοντέλο συγκρίνονται παρακάτω με γνωστά αποτελέσματα νευροαπεικόνισης για την κινητική απεικόνιση.



Σχήμα 5.1 Η προβολή των τιμών SHAP σε εγκεφαλικό φλοιό

Αναφέρουμε τα ευρήματα ανά κλάση, επισημαίνοντας κατά πόσο οι περιοχές του φλοιού με τη μεγαλύτερη επιρροή ευθυγραμμίζονται με τις αναμενόμενες σωματοτοπικές κινητικές αναπαραστάσεις και τα γνωστά μοτίβα πλευροποίησης. Παρέχονται περιλήψεις πινάκων για κάθε κλάση, όπου παρατίθενται οι περιοχές με την υψηλότερη συνεισφορά (θετική και αρνητική Επιπλέον, χρησιμοποιήσαμε το προαναφερθέν προσαρμοσμένο εγκεφαλικό διάγραμμα SHAP που μετατρέπει τις τιμές SHAP σε πίνακες σε έναν πιο διαισθητικό τρόπο κατανόησης των εγκεφαλικών περιοχών που είναι «σχηματικά» σημαντικές. Ο εγκέφαλος ακολουθεί την απλή προβολή κάθε τιμής SHAP στο brain plot, όπως φαίνεται στο Σχήμα 5.4.

#### 5.1.1 Απεικόνιση νοερής κίνησης αριστερού χεριού (Left Hand Imagery)

Για τη νοερή κίνηση του αριστερού χεριού, η ανάλυση ερμηνευσιμότητας του μοντέλου ανέδειξε τον αισθητικοκινητικό φλοιό του δεξιού ημισφαιρίου ως την περιοχη με την περισσότερη πληροφορία για το μοντέλο (σε σχέση με την πρόβλεψη της «αριστερής» κλάσης). Ειδικότερα, ο **Right Precentral Gyrus** (δεξιός πρωτοταγής κινητικός φλοιός, *M1*) εμφάνισε τη μεγαλύτερη θετική τιμή SHAP, ακολουθούμενος από τον **Right Postcentral Gyrus** (πρωτοταγής σωματοαισθητικός φλοιός, *S1*). Οι περιοχές αυτές είναι ετερόπλευρες (contralateral) ως προς το νοούμενο —αριστερό— χέρι και, έτσι, εναρμονίζονται με την οργάνωση του κινητικού homunculus (όπως παρουσιάστηκε από τον Penfield [196]. Τόσο ο συνολικός μέσος όρος SHAP όσο και η ανάλυση εστιασμένη στον χρόνο (στο στατιστικά μέγιστο παράθυρο νοερής κίνησης) κατέδειξαν αυτήν τη δεξιόπλευρη αισθητικοκινητική κυριαρχία. Το εύρημα υποστηρίζεται ισχυρά από προηγούμενες μελέτες fMRI: η νοερή κίνηση χεριών ενεργοποιεί σταθερά την ετερόπλευρη περιοχή χεριού του *M1* και τις παρακείμενες προκινητικές/σωματοαισθητικές περιοχές [197]. Συνεπώς, η ταύτιση των **Right M1/S1** από το μοντέλο ως βασικών συντελευρο έλεγχο των κινήσεων χεριού.

Αντιθέτως, οι κινητικές περιοχές του *αριστερού* ημισφαιρίου (ομόπλευρες ως προς το νοούμενο χέρι) εμφάνισαν τις πιο αρνητικές τιμές SHAP για την κλάση «αριστερό χέρι». Ιδίως, ο Left Precentral Gyrus είχε ισχυρή αρνητική συνεισφορά — πράγμα που σημαίνει

ότι, εάν ο αριστερός M1 ενεργοποιείται, το μοντέλο είναι λιγότερο πιθανό να προβλέψει «αριστερό χέρι».



Σχήμα 5.2 Οι τιμές SHAP για κάθε μια από τις κλάσεις, (a) "αριστερό", (b) "δεζί", (c) "πόδια" και (d) "γλώσσα"

Το εύρημα αυτό ανταποκρίνεται στις νευροφυσιολογικές προσδοκίες: κατά τη μονόπλευρη νοερή κίνηση, ο ομόπλευρος κινητικός φλοιός συνήθως εμφανίζει μειωμένη ή ακόμη και ανασταλτική δραστηριότητα, ενώ ο ετερόπλευρος είναι ενεργός [198], [199]. Με άλλα λόγια, ισχυρή ενεργοποίηση του αριστερού *M1* αποτελεί χαρακτηριστικό νοερής κίνησης δεξιού χεριού<sup>•</sup> επομένως, στον ταξινομητή αριστερού χεριού λειτουργεί ως «αντι-τεκμήριο». Το μοντέλο ανέθεσε επίσης αρνητική βαρύτητα στον **Left Postcentral Gyrus**, συμβαδίζοντας με την άποψη ότι η ομόπλευρη σωματοαισθητική εμπλοκή δεν είναι εξέχουσα στη σωστή νοερή κίνηση αριστερού χεριού. Επιπλέον, ορισμένες περιοχές μη σχετιζόμενες με το χέρι εμφανίσθηκαν ως ήπια αρνητικές συνεισφορές — π.χ. το μοντέλο μείωσε το βάρος στις έσω μετωπιαίες (paracentral) περιοχές που συνδέονται με αναπαράσταση ποδιού/κνήμης<sup>•</sup> λογικό, αφού ενεργοποίηση στη ζώνη ποδιού θα υποδήλωνε ότι ο συμμετέχων εκτελεί νοερή κίνηση ποδιού αντί για αριστερό χέρι. Συνολικά, το προφίλ SHAP για αριστερό χέρι (δεξιόπλευρα pro-features, αριστερό χέρι των δικτύων νοερής κίνησης [199].

| Εγκεφαλική<br>περιοχή<br>(Ημισφαίριο)          | Συνεισφορά<br>SHAP                   | Γνωστή συνάφεια για νοερή κίνηση αριστερού<br>χεριού   |
|--|--------------------------------------|--|
| <b>Right Precentral</b><br><b>Gyrus (R M1)</b> | Μέγιστη θετική<br>(pro-class)        | Ετερόπλευρη πρωτοταγής κινητική περιοχή<br>χεριού· ενεργοποιείται σε νοερή κίνηση<br>αριστερού χεριού [197]  |
| Right Postcentral<br>Gyrus (R S1)              | Υψηλή θετική                         | Ετερόπλευρη σωματοαισθητική περιοχή χεριού·<br>συμπαρασύρεται με <i>Μ1</i> προσφέροντας<br>κιναισθητική ανατροφοδότηση [198]                                 |
| Left Precentral<br>Gyrus (L M1)                | Μέγιστη<br>αρνητική (anti-<br>class) | Ομόπλευρος κινητικός φλοιός· η ενεργοποίησή<br>του δηλώνει νοερή κίνηση δεξιού χεριού (οι<br>κινητικοί χάρτες είναι κυρίως ετερόπλευροι)<br>[199]            |
| Left Postcentral<br>Gyrus (L S1)               | Υψηλή<br>αρνητική                    | Ομόπλευρη σωματοαισθητική περιοχή· δεν<br>δραστηριοποιείται έντονα σε αριστερό ΜΙ·<br>ισχυρή δραστηριότητα θα υποδήλωνε άλλο έργο<br>(π.χ. ΜΙ δεξιού χεριού) |

Πίνακας 4 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση αριστερού χεριού και η γνωστή συνάφειά τους

#### 5.2.2 Απεικόνιση νοερής κίνησης δεξιού χεριού (Right Hand Imagery)

Το μοτίβο για τη νοερή κίνηση δεξιού χεριού αποτελεί ουσιαστικά το κατοπτρικό είδωλο της περίπτωσης του αριστερού χεριού. Η ανάλυση SHAP ανέδειξε τις αισθητικο-κινητικές περιοχές του αριστερού ημισφαιρίου ως τους σημαντικότερους θετικούς συντελεστές για την κλάση «δεξιού χεριού». Ο Left Precentral Gyrus (αριστερός *M1*) εμφάνισε τη μέγιστη θετική τιμή SHAP, ακολουθούμενος από τον Left Postcentral Gyrus (αριστερός *S1*). Η ενεργοποίηση στις αριστερές πρωτοταγείς κινητικές/αισθητικές περιοχές ωθεί έντονα το μοντέλο προς την πρόβλεψη «δεξί χέρι», κάτι που συμφωνεί πλήρως με τη νευροεπιστήμη: οι κινήσεις (ή η νοερή κίνηση) του δεξιού χεριού ελέγχονται κυρίως από τον αριστερό κινητικό φλοιό [197]. Μελέτες fMRI επιβεβαιώνουν ότι η φαντασιακή κίνηση δεξιού χεριού ενεργοποιεί την περιοχή χεριού στον αριστερό *M1* καθώς και προκινητικές / συμπληρωματικές κινητικές περιοχές του ίδιου ημισφαιρίου [200]. Η έξοδος του μοντέλου μας ευθυγραμμίζεται με αυτήν την ετερόπλευρη αντιστοίχιση—αναγνωρίζει την αριστερή φλοιϊκή δραστηριότητα ως σήμα-κατατεθέν της νοερής κίνησης δεξιού χεριού.

Αντίστροφα, οι κινητικές περιοχές του δεξιού ημισφαιρίου λειτούργησαν ως οι πιο αρνητικοί παράγοντες SHAP για την κλάση «δεξιού χεριού». Ο Right Precentral Gyrus είχε έντονη αρνητική συνεισφορά, δηλώνοντας ότι εάν ο δεξιός M1 εμφανίσει δραστηριότητα, το μοντέλο είναι λιγότερο πιθανό να προβλέψει «δεξί χέρι». Φυσιολογικά, ισχυρή ενεργοποίηση δεξιού M1 υποδηλώνει μάλλον νοερή κίνηση **αριστερού** χεριού, αντικατοπτρίζοντας και πάλι το ετερόπλευρο πρότυπο [199]. Ανάλογο αρνητικό βάρος έφερε και ο **Right Postcentral Gyrus**. Ουσιαστικά, το μοντέλο έμαθε να διακρίνει νοερή κίνηση δεξιού από αριστερού χεριού βάσει της πλευρίωσης της αισθητικοκινητικής συμμετοχής: ενεργό αριστερό sensorimotor (και σχετικά ανενεργό δεξί)  $\rightarrow$  «δεξί χέρι»· το αντίστροφο  $\rightarrow$  «αριστερο χέρι». Αυτό συμφωνεί με δεκαετίες ευρημάτων πάνω στη μου-ρυθμική αποσυγχρονισμένη τοπογραφία ΕΕG, όπου η νοερή κίνηση δεξιού χεριού προκαλεί εμφανές ERD στην κεντρική

αριστερή περιοχή του κρανίου (ηλεκτρόδιο C3), ενώ η νοερή κίνηση αριστερού χεριού εμφανίζει ERD στη δεξιά κεντρική περιοχή (C4) [201].



Σχήμα 5.3 Απεικόνιση Gradient SHAP στον εγκέφαλο για τη νοερή κίνηση δεξιού χεριού, στον χρόνο 0,348 s.

Με τη χρήση χαρακτηριστικών σε επίπεδο πηγής, το μοντέλο επιβεβαιώνει ότι αυτά τα μοτίβα αισθητήρα πράγματι πηγάζουν από τους αριστερούς και δεξιούς Precentral Gyri αντί για ασαφείς μικτές πηγές.

Πίνακας 5 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση δεξιού χεριού και η γνωστή συνάφειά τους

| Εγκεφαλική<br>περιοχή<br>(Ημισφ.) | Συνεισφορά<br>SHAP | Γνωστή συνάφεια για νοερή κίνηση δεξιού<br>χεριού                                       |
|-----------------------------------|--------------------|---|
| Left Precentral                   | Μέγιστη θετική     | Ετερόπλευρη πρωτοταγής κινητική περιοχή   |
| Gyrus (L M1)                      | (pro-class)        | δεξιού χεριού· ενεργοποιείται σε νοερή κίνηση<br>δεξιού χεριού [197], [200]             |
| Left Postcentral<br>Gvrus (L S1)  | Υψηλή θετική       | Ετερόπλευρη σωματοαισθητική περιοχή δεξιού<br>γεριού· εμπλέκεται στη κιναισθητική νοερή |
|                                   |                    | κίνηση δεξιού χεριού [201], [202], [203]  |
| <b>Right Precentral</b>           | Μέγιστη            | Ομόπλευρος κινητικός φλοιός· η ενεργοποίησή   |
| Gyrus (R M1)                      | αρνητική (anti-    | του ασύμβατη ή/και κατασταλμένη στη νοερή   |
|                                   | class)             | κίνηση δεξιού χεριού [199]  |
| Right                             | Υψηλή αρνητική     | Ομόπλευρη σωματοαισθητική περιοχή· δεν  |
| Postcentral                       |                    | χαρακτηρίζει αυθεντική ΜΙ δεξιού χεριού·  |
| Gyrus (R S1)                      |                    | δραστηριότητα εδώ υποδηλώνει άλλη κλάση (π.χ.<br>αριστερό χέρι)                         |

Πέραν των πρωτοταγών αισθητικοκινητικών περιοχών, το μοντέλο απέδωσε μέτρια θετική σημασία SHAP στον Left Superior Frontal Gyrus κατά τη νοερή κίνηση δεξιού χεριού, ιδίως στην ανάλυση χρονικού παραθύρου. Η περιοχή αυτή αντιστοιχεί πιθανώς στον αριστερό ραχιαίο προκινητικό φλοιό ή/και στη συμπληρωματική κινητική περιοχή (SMA). Προηγούμενες απεικονιστικές μελέτες δείχνουν ότι η νοερή κίνηση χεριού δεσμεύει όχι μόνο M1 αλλά και προκινητικός/μέσες κινητικές περιοχές στο ετερόπλευρο ημισφαίριο [197], [202]. Επομένως, η συμβολή αριστερού premotor / SMA στη νοερή κίνηση δεξιού χεριού συνάδει με τα γνωστά δίκτυα σχεδιασμού και απεικόνισης κίνησης [201], [203]. Αντίστοιχα, ο Right Superior Frontal Gyrus είχε ελαφρά αρνητική τιμή SHAP, γεγονός που πάλι υποδηλώνει ότι η ενεργοποίησή του θα παρέπεμπε σε νοερή κίνηση αριστερού χεριού. Συνοπτικά, οι σημαντικές περιοχές που προκύπτουν από το SHAP για το δεξί χέρι είναι πλήρως συνεπείς με την αναμενόμενη ετερόπλευρη κινητική εμπλοκή. Ο Πίνακας 5 παρουσιάζει τις κορυφαίες περιοχές και τη βιβλιογραφική τους τεκμηρίωση.

#### 5.2.3 Απεικόνιση νοερής κίνησης ποδιών (Foot Imagery)

Για την κλάση «πόδια»—όπου οι συμμετέχοντες φαντάζονταν κίνηση αμφότερων των ποδιών—οι σημαντικότερες θετικές τιμές SHAP αντιστοιχούσαν σε έσω (medial) φλοιϊκές περιοχές που σχετίζονται με τον έλεγχο των κάτω άκρων. Ειδικότερα, ο Paracentral Lobule και στα δύο ημισφαίρια εμφάνισε τις υψηλότερες θετικές τιμές. Ο λοβός αυτός, στο έσω τοίχωμα κάθε ημισφαιρίου, περιλαμβάνει τη ζώνη ποδιού/κνήμης των πρωτοταγών κινητικών και σωματοαισθητικών φλοιών—το «κορυφαίο» τμήμα του κινητικού homunculus. Η εξάρτηση του μοντέλου και από τις δύο (αριστερή & δεξιά) paracentral περιοχές συνάδει με τη νοερή κίνηση και των δύο ποδιών, η οποία ενεργοποιεί τις αναπαραστάσεις ποδιού και στα δύο ημισφαίρια. Το εύρημα συμφωνεί με δεδομένα fMRI/MEG που δείχνουν ότι η νοερή κίνηση ποδιού ενεργοποιεί τις άνω-έσω κινητικές περιοχές· λ.χ. φανταστική κίνηση δεξιού ποδιού προκαλεί δραστηριότητα στον ετερόπλευρο (αριστερό) Paracentral Lobule και τη συμπληρωματική κινητική περιοχή (SMA) [197], [204], ενώ η φανταστική κίνηση και των δύο ποδιών οδηγεί σε διμερή ενεργοποίηση. Πράγματι, ο Ehrsson και συνεργάτες ανέφεραν ότι η νοερή κίνηση ποδιού εμπλέκει ειδικά τις «ζώνες ποδιού» του SMA και του πρωτοταγούς κινητικού φλοιού [197]. Τα δικά μας αποτελέσματα το αντικατοπτρίζουν: η «global» σημασία SHAP έδειξε τον διμερή Paracentral/SMA ως κυρίαρχα χαρακτηριστικά, ενώ η χρονικά εστιασμένη ανάλυση τόνισε ακόμη περισσότερο αυτές τις περιοχές κατά το μέγιστο της νοερής κίνησης.

Επιπλέον, το μοντέλο εντόπισε θετική συνεισφορά στη medial Superior Frontal Gyrus (πιθανή επικάλυψη με το ίδιο το SMA), γεγονός που συμβαδίζει με τον καθοριστικό ρόλο του SMA στη νοερή κίνηση και τον συντονισμό των κάτω άκρων [197]. Παρατηρήθηκε επίσης αλλά χαμηλότερα στην κατάταξη—θετική βαρύτητα στο Precuneus / οπίσθιο βρεγματικό λοβό· περιοχή που έχει συσχετιστεί με κινητική απεικόνιση και οπτικο-χωρική αναπαράσταση στάσης ή κίνησης στο χώρο [205].

Ως προς τις αρνητικές συνεισφορές, ξεχώρισαν οι πλάγιες φλοιϊκές περιοχές που σχετίζονται με κινήσεις χεριού και προσώπου. Ο μοντέλο απέδωσε αρνητικές τιμές SHAP στον αριστερό και δεξιό **Precentral Gyrus** (πλάγιες περιοχές—hand area)· δηλαδή, έντονη δραστηριότητα σε ζώνες χεριού μειώνει την πιθανότητα πρόβλεψης «foot». Αυτό αντικατοπτρίζει την αρχή της σωματοτοπικής αποκλειστικότητας: ενεργοποίηση στο hand motor cortex υποδηλώνει καθαρά — ότι το έργο αφορά χέρι κι όχι πόδι. Ομοίως, οι κατώτερες μετωπιαίες περιοχές που αντιστοιχούν σε κίνηση στόματος/γλώσσας έφεραν αρνητικές τιμές· λογικά, η ενεργοποίηση ζώνης γλώσσας θα παρέπεμπε σε έργο «tongue» κι όχι «foot». Συνολικά, τα χαρακτηριστικά που έμαθε το μοντέλο ενισχύουν τη διακριτότητα των αναπαραστάσεων ποδιού: αναζητεί ενεργοποίηση των **έσω** κινητικών περιοχών (θετικό SHAP) και απουσία δραστηριότητας στον πλάγιο κινητικό φλοιό (που, αν υπάρχει, γέρνει το μοντέλο προς κλάσεις χεριού ή γλώσσας). Η διάκριση αυτή αντανακλά ακριβώς την οργάνωση του homunculus όπως τεκμηριώνεται από fMRI και ηλεκτροδιεγερτικούς χάρτες: η νοερή κίνηση ποδιού ενεργοποιεί τον άνω-έσω κινητικό φλοιό, όχι τις πλάγιες κινητικές ζώνες [197]. Στις αποδόσεις MI αριστερού χεριού παρατηρήθηκε αξιοσημείωτη χρονική μεταβολή: νωρίς στο δοκίμιο (~0–0,5 s από την έναρξη), ισχυρές θετικές τιμές SHAP εμφανίζονται στον **Left Superior Parietal** και **Paracentral**, ενώ ο **Left Postcentral Gyrus** παρουσιάζει αρνητικά «αιχμηρά» SHAP· δηλαδή, η έλλειψη (ή σχετική απενεργοποίηση) του αριστερού *Postcentral* αποτελεί ένδειξη νοερής κίνησης αριστερού χεριού (καθώς η ενεργοποίησή του αντιστοιχεί σε ΜΙ δεξιού χεριού).

| Εγκεφαλική<br>περιοχή (Ημισφ.)                          | Συνεισφορά<br>SHAP                 | Γνωστή συνάφεια για Foot Imagery   |
|---|------------------------------------|--|
| Left Paracentral<br>Lobule (medial L<br>M1/S1)          | Μέγιστη<br>θετική (pro-<br>class)  | Έσω πρωτοταγής κινητική/αισθητική ζώνη<br>ποδιού· ενεργοποιείται σε φαντασιακή κίνηση<br>ποδιού (ετερόπλευρα) [197], [204]· σε διμερή ΜΙ<br>ποδιών ενεργοποιούνται αμφότερες πλευρές |
| Right Paracentral<br>Lobule (medial R<br>M1/S1)         | Υψηλή θετική                       | Έσω κινητική ζώνη δεξιού ποδιού·<br>ενεργοποιείται σε ΜΙ δεξιού ποδιού· υποστηρίζει<br>διμερή ενεργοποίηση στη ΜΙ και των δύο ποδιών<br>[197], [204]                                 |
| Superior Frontal<br>Gyrus (medial<br>aspect / SMA)      | Θετική                             | Συμπληρωματική κινητική περιοχή στο έσω<br>τοίχωμα· εμπλέκεται στον σχεδιασμό/απεικόνιση<br>κινήσεων κάτω άκρων· κρίσιμη για διμερή<br>συντονισμό [205]                              |
| Left Precentral<br>Gyrus (lateral L<br>M1 – hand area)  | Υψηλή<br>αρνητική (anti-<br>class) | Πλάγιος κινητικός φλοιός χεριού· η<br>ενεργοποίηση αντιφάσκει με ΜΙ ποδιών και<br>δείχνει έργο χεριού [197]  |
| Right Precentral<br>Gyrus (lateral R<br>M1 – hand area) | Υψηλή<br>αρνητική                  | Πλάγιος κινητικός φλοιός χεριού στο δεξί<br>ημισφαίριο· δραστηριότητα εδώ δηλώνει έργο<br>χεριού, μειώνοντας την πιθανότητα «foot»   |
| Inferior Frontal<br>(Opercular)<br>Cortex (Left)        | Αρνητική<br>(αξιόλογη)             | Περιοχή Broca/κοιλιακός προκινητικός (ζώνη<br>στόματος/γλώσσας)· μη αναμενόμενη σε ΜΙ<br>ποδιών· δραστηριότητα υποδηλώνει έργο<br>γλώσσας  |

Πίνακας 6 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση ποδιών και η γνωστή συνάφειά τους

#### 5.2.4 Απεικόνιση νοερής κίνησης γλώσσας (Tongue Imagery)

Η κλάση νοερής κίνησης γλώσσας εμφάνισε ένα ιδιαίτερο μοτίβο σημαντικών περιοχών, συμβατό με το γεγονός ότι οι στοματοπροσωπικές αναπαραστάσεις καταλαμβάνουν τα πιο πλάγια τμήματα του αισθητικοκινητικού φλοιού και συνήθως ενεργοποιούν αμφίπλευρα δίκτυα. Το μοντέλο αναγνώρισε τους lateral Precentral Gyri αμφοτέρων των ημισφαιρίων ως κορυφαίους θετικούς συντελεστές. Συγκεκριμένα, ο Right Precentral Gyrus (κατώτερο

τμήμα) έλαβε τη μέγιστη θετική τιμή SHAP, ακολουθούμενος στενά από τον Left Precentral Gyrus (κατώτερο τμήμα). Οι περιοχές αυτές αντιστοιχούν στη ζώνη προσώπου/γλώσσας του πρωτοταγούς κινητικού φλοιού σε δεξί και αριστερό ημισφαίριο, αντίστοιχα. Η αμφίπλευρη συνεισφορά τους συμφωνεί με νευροαπεικονιστικά ευρήματα ότι η κίνηση ή η νοερή κίνηση γλώσσας/χειλιών ενεργοποιεί τον κινητικό φλοιό και στις δύο πλευρές. Σε αντίθεση με τις κινήσεις άκρων (κυρίως ετερόπλευρες), ο έλεγχος της γλώσσας είναι περισσότερο διπλής πλευράς λόγω της μέσης γραμμής των μυών και της διπλής εννεύρωσης. Μελέτες fMRI επιβεβαιώνουν ότι η φανταστική κίνηση γλώσσας ενεργοποιεί διμερώς τη γλωσσική περιοχή του *M1* και τον πλάγιο προκινητικό φλοιό [197], [204]. Η ερμηνευσιμότητα του μοντέλου αντικατοπτρίζει ακριβώς αυτό: θεωρεί τη δραστηριότητα είτε στον αριστερό είτε στον δεξιό κατώτερο *Precentral* ως υποστηρικτική ένδειξη γλωσσικής νοερής κίνησης. Ενδιαφέρον παρουσιάζει η ανάλυση χρονικού παραθύρου, όπου παρατηρήθηκε συμμετρική αύξηση σημαντικότητας στους δύο *Precentral* κατά τη φάση της νοερής κίνησης, υποδεικνύοντας ότι το μοντέλο ανιχνεύει το αμφίπλευρο πρότυπο ως σήμα-υπογραφή για την κλάση «tongue» (σε αντίθεση με τα μονοπλευρικά πρότυπα των κλάσεων χεριού).

Πέρα από τις πρωτοταγείς κινητικές περιοχές, ο Left Inferior Frontal Gyrus (ιδίως το τμήμα opercular—περιοχή Broca) αποτέλεσε επιπλέον θετική συνεισφορά. Αν και δεν είναι πρωτοταγής κινητική περιοχή γλώσσας, η περιοχή Broca συνδέεται λειτουργικά με τον στοματοπροσωπικό έλεγχο και τον σχεδιασμό ομιλίας. Στο μοντέλο μας, η εμπλοκή της Broca έλαβε σημαντική θετική τιμή SHAP, υποδηλώνοντας ότι, όταν ανιχνεύεται ενεργοποίηση εκεί, η πρόβλεψη κλίνει προς «tongue imagery». Νευροβιολογικά, αυτό είναι συνεισφόρον και εναρμονίζεται με παρατηρήσεις ότι η νοερή κίνηση γλώσσας μπορεί να συνενεργοποιεί περιοχές λόγου και ανώτερου κινητικού σχεδιασμού [206]. Η γλώσσα συνδέεται στενά με την ομιλία, και η Broca—κλασική γλωσσική περιοχή—ενδέχεται να στρατολογείται ακόμη και σε σιωπηλή νοητική κίνηση της γλώσσας [205]. Μελέτες συνδεσιμότητας δείχνουν ότι η γλωσσική ΜΙ εμπλέκει ένα βρεγματο-μετωπιαίο δίκτυο που περιλαμβάνει premotor κι ευρύτερα περιοχές Broca, πιθανώς αντικατοπτρίζοντας εσωτερικό λόγο ή τον εξελικτικό ρόλο της γλώσσας στην ομιλία [205]. Η παρουσία της Broca στις εξηγήσεις του μοντέλου δείχνει πώς η βαθιά μάθηση μπορεί να αποκαλύψει τη συμμετοχή ανώτερων περιοχών πέραν του κύριου αισθητικοκινητικού φλοιού.

Οι αρνητικές τιμές SHAP για την κλάση γλώσσας προήλθαν κυρίως από περιοχές που σχετίζονται με κινήσεις άκρων. Λ.χ. οι **Paracentral Lobules** (ζώνες ποδιού) και στα δύο ημισφαίρια είχαν ισχυρές αρνητικές τιμές—αν το μοντέλο ανιχνεύσει δραστηριότητα στον έσω κινητικό φλοιό (όπως θα γινόταν σε foot imagery), καταστέλλει την πιθανότητα «tongue», κάτι λογικό, καθώς η ενεργοποίηση ζώνης ποδιού είναι ασύμβατη με γλωσσικό έργο. Ομοίως, τα hand-area τμήματα του *Precentral* έλαβαν αρνητικό βάρος: έντονη δραστηριότητα στον «χειρομορφικό» θόλο του *M1* υποδηλώνει MI χεριού, όχι γλώσσας. Έτσι, το μοντέλο έμαθε ότι η **απουσία** εξειδικευμένων μοτίβων άκρων είναι προϋπόθεση για ταξινόμηση γλώσσας. Αυτό συνάδει με τη σωματοτοπική οργάνωση: το έργο «tongue» πρέπει να ενεργοποιεί μοναδικά πλάγιες μετωπιαίες περιοχές, ενώ δραστηριότητα σε ραχιαίες ή έσω ζώνες δείχνει άλλες κλάσεις. Αξίζει να σημειωθεί ότι η ΜΙ γλώσσας μπορεί να παράγει ελάχιστη απόκριση σε κεντρικά ηλεκτρόδια που αλληλεπικαλύπτονται με hand areas· το μοντέλο το χειρίζεται αντιμετωπίζοντας αυτές τις κεντρικές περιοχές ως anti-features, διασφαλίζοντας πως η κλάση «tongue» επιφυλάσσεται για δοκιμές με χαρακτηριστική αμφίπλευρη πλάγια μετωπιαία ενεργοποίηση.

Συνολικά, η ανάλυση SHAP για τη γλωσσική MI υπογραμμίζει ένα αμφίπλευρο πρότυπο φλοιϊκής ενεργοποίησης (και των δύο κινητικών φλοιών) με επέκταση σε γλωσσικάσχετιζόμενες μετωπιαίες περιοχές. Αυτό συμφωνεί άριστα με τη βιβλιογραφία fMRI, η οποία αναφέρει ότι η νοερή κίνηση γλώσσας/χειλιών εμπλέκει διμερώς M1 και τον premotor φλοιό, συχνά πιο συμμετρικά από ό,τι η MI άκρων [197]. Επιπλέον, το απροσδόκητο εύρημα συμμετοχής της Broca στηρίζεται από την υπόθεση ότι η νοερή γλώσσα ενδέχεται να ενεργοποιεί δίκτυα ομιλίας/κινητικής απεικόνισης [205]. Ο Πίνακας 7 συνοψίζει τις κορυφαίες περιοχές για την κλάση γλώσσας και την ερμηνεία τους με βάση προηγούμενες μελέτες.

| Εγκεφαλική<br>περιοχή (Ημισφ.)                    | Συνεισφορά<br>SHAP                   | Γνωστή συνάφεια για Tongue Imagery   |
|---|--------------------------------------|--|
| Right Precentral<br>Gyrus (R M1 –<br>tongue area) | Μέγιστη θετική<br>(pro-class)        | Πλάγιος πρωτοταγής κινητικός φλοιός (περιοχή<br>προσώπου/γλώσσας) ετερόπλευρα της αριστερής<br>πλευράς γλώσσας· η ΜΙ γλώσσας ενεργοποιεί<br>συχνά διμερώς αυτήν την περιοχή [197], [204] |
| Left Precentral<br>Gyrus (L M1 –<br>tongue area)  | Υψηλή θετική                         | Πλάγιος πρωτοταγής κινητικός φλοιός στο<br>αριστερό ημισφαίριο· η νοερή κίνηση γλώσσας<br>ενεργοποιεί τα γλωσσικά υπο-μοτίβα του <i>M1</i> και<br>στις δύο πλευρές [197], [204]          |
| Left Inferior<br>Frontal Gyrus<br>(Broca's Area)  | Θετική<br>(αξιόλογη)                 | Περιοχή συσχέτισης κινητικού λόγου· συνδέεται<br>με τον γλωσσικό κινητικό φλοιό [205]· η<br>ενεργοποίησή της υποδηλώνει συν-στράτευση<br>γλωσσικού σχεδιασμού                            |
| Left Postcentral<br>Gyrus (L S1 –<br>face area)   | Μέτρια θετική                        | Πλάγιος σωματοαισθητικός φλοιός<br>(αναπαράσταση στόματος/γλώσσας)· πιθανή<br>εμπλοκή μέσω κιναισθητικής ανατροφοδότησης   |
| Left Paracentral<br>Lobule (L leg<br>area)        | Μέγιστη<br>αρνητική (anti-<br>class) | Έσω κινητικός φλοιός (ζώνη ποδιού)·<br>ενεργοποίηση εδώ δηλώνει ΜΙ ποδιού, όχι<br>γλώσσας [197]  |
| Right Paracentral<br>Lobule (R leg<br>area)       | Υψηλή<br>αρνητική                    | Έσω κινητικός φλοιός δεξιού ποδιού·<br>δραστηριότητα κυρίως σε foot MI, άρα μειώνει<br>την πιθανότητα «tongue»   |
| Left Precentral<br>(dorsal – hand<br>area)        | Αρνητική                             | Ραχιαίος <i>M1</i> (ζώνη χεριού) στο αριστερό<br>ημισφαίριο· ενεργοποίηση υποδηλώνει MI χεριού<br>(π.χ. δεξιού), αντιτάσσεται στην κλάση γλώσσας   |

Πίνακας 7 Κορυφαίες περιοχές (SHAP) για τη νοερή κίνηση γλώσσας και η γνωστή συνάφειά τους

## 5.5 Συζήτηση

#### Συνέπεια με τη νευροαπεικόνιση και τη σωματοτοπική οργάνωση.

Τα παραπάνω αποτελέσματα καταδεικνύουν υψηλό βαθμό συμφωνίας μεταξύ των εξόδων του μοντέλου βαθιάς μάθησης και της εδραιωμένης νευροεπιστημονικής γνώσης για τη νοερή κίνηση. Οι περιοχές-κλειδιά που ανέδειξε το SHAP για κάθε κλάση προβάλλονται εύστοχα πάνω στη προβλεπόμενη σωματοτοπική διάταξη του κινητικού φλοιού, αναπαράγοντας ουσιαστικά έναν λειτουργικό «homunculus» σύμφωνο με τα πορίσματα του Penfield και

μεταγενέστερων μελετών fMRI. Ειδικότερα, η νοερή κίνηση χεριού συνδέθηκε κυρίως με ετερόπλευρες προκεντρικές περιοχές (hand area του M1), η κίνηση ποδιού με τους διμερείς παρακεντρικούς λοβούς (leg areas) και η κίνηση γλώσσας με διμερείς κατώτερες προκεντρικές ζώνες (face/tongue areas). Την αντιστοίχιση αυτή επιβεβαιώνουν πολυάριθμες δημοσιευμένες μελέτες ο Ehrsson et al. έδειξε ρητώς ότι η νοερή κίνηση δακτύλων, δακτύλων ποδιού και γλώσσας ενεργοποιεί τα αντίστοιχα τμήματα του κινητικού φλοιού [197]. Η καθολική ανάλυση SHAP του μοντέλου επανεντόπισε πρακτικά αυτές τις χαρτογραφήσεις, επικυρώνοντας ότι η απόφαση του CNN βασίζεται σε γνήσια νευροανατομικά πρότυπα και όχι σε τυχαία χαρακτηριστικά EEG. Οι θετικές περιοχές SHAP (pro-class) αντιστοιχούν στα σημεία όπου αναμένεται φλοιϊκή ενεργοποίηση για τη συγκεκριμένη νοερή κίνηση, ενώ οι αρνητικές περιογές SHAP αντιστοιγούν σε ζώνες που ενεργοποιούνται σε ανταγωνιστικές κλάσεις – εύρημα λογικό, αν ληφθεί υπόψη ο αμοιβαίως αποκλειόμενος χαρακτήρας των επιμέρους εργασιών. Η συνέπεια αυτή είναι κρίσιμη: δείχνει ότι ο ταξινομητής εκμεταλλεύεται σήματα με πραγματική φυσιολογική βάση (π.χ. αποσυγχρονισμό ρυθμού μ σε κατάλληλες περιοχές) και δεν «κλέβει» μέσω artefacts ή θορύβου.

#### Καταγραφή φαινομένων πλευρίωσης.

Το μοντέλο κατέγραψε επίσης τα γνωστά αποτελέσματα πλευρίωσης. Tanto η νοερή κίνηση αριστερού όσο και δεξιού γεριού εμφάνισε ετερόπλευρη κυριαργία στα ερμηνευτικά χαρακτηριστικά, αντανακλώντας τον διαημισφαιρικό έλεγχο των άκρων [199]. Το εύρημα συνάδει με κλασικά δεδομένα EEG και fNIRS· λ.χ. η νοερή κίνηση δεξιού χεριού προκαλεί εντονότερο ERD πάνω από τον αριστερό κινητικό φλοιό, ενώ η αντίστοιχη αριστερού χεριού πάνω από τον δεξιό [201]. Η πηγαιοτοποθετημένη προσέγγιση μας επέτρεψε να συνδέσουμε αυτές τις παρατηρήσεις επιπέδου αισθητήρων με συγκεκριμένα φλοιϊκά ROI (αριστερός ή δεξιός προκεντρικός έλικας), ενισχύοντας την εμπιστοσύνη ότι οι εσωτερικές αναπαραστάσεις του μοντέλου αντιστοιχούν σε αληθή φλοιϊκή φυσιολογία. Παρατηρήθηκε μία λεπτή διαφοροποίηση: ο γάρτης SHAP για τη νοερή κίνηση του αριστερού (μη κυρίαρχου) χεριού ήταν ελαφρώς λιγότερο μονοπλεύρος από εκείνον για το δεξί (κυρίαρχο) χέρι, εύρημα που συμφωνεί με αναφορές ότι δεξιόχειρες συμμετέχοντες εμφανίζουν ενίοτε πιο διμερή ενεργοποίηση όταν φαντάζονται κινήσεις του αριστερού γεριού [198]. Ακόμη κι αυτές οι μικρές απογρώσεις καταδεικνύουν ότι τα γαρακτηριστικά που έμαθε το μοντέλο είναι ευαίσθητα σε αυθεντικά νευρωνικά πρότυπα, συμπεριλαμβανομένων εκείνων που επηρεάζονται από την προτίμηση γειρός και την ημισφαιρική κυριαρχία.

#### Κλινική και επιστημονική σημασία

Η σύγκλιση των ερμηνεύσιμων εξόδων του μοντέλου μας με περιοχές που έχουν επιβεβαιωθεί μέσω fMRI έχει ουσιώδεις συνέπειες τόσο για εφαρμογές BCI όσο και για τη νευροεπιστήμη. Σε κλινικά περιβάλλοντα (π.χ. συστήματα αποκατάστασης μετά από εγκεφαλικό), είναι ενθαρρυντικό ότι το μοντέλο «θεμελιώνει» τις ταξινομήσεις του στις αναμενόμενες κινητικές περιοχές, υποδηλώνοντας ότι η διεπαφή εμπλέκει πράγματι τα επιδιωκόμενα νευρωνικά κυκλώματα. Για παράδειγμα, αν ένα BCI αποκατάστασης στοχεύει στην ενεργοποίηση του ipsilesional κινητικού φλοιού μέσω νοερής κίνησης, ένα ερμηνεύσιμο μοντέλο μπορεί να επαληθεύσει ότι ο ασθενής πράγματι στρατολογεί τη συγκεκριμένη περιοχή (υψηλή τιμή SHAP εκεί) όταν το BCI ανιχνεύει προσπάθεια [197]. Αντιστρόφως,

εάν το μοντέλο υποδείκνυε μη αναμενόμενες περιοχές, ο κλινικός θα ανησυχούσε ότι ο ασθενής εφαρμόζει υποβέλτιστη στρατηγική (π.χ. οπτική φαντασία αντί κινητικοκιναισθητικής). Κατά συνέπεια, τα αποτελέσματά μας ενισχύουν την εμπιστοσύνη πως οι ταξινομητές βαθιάς μάθησης μπορούν να ερμηνευθούν νευροανατομικά· βήμα καίριο για κλινική υιοθέτηση. Γεφυρώνεται έτσι το χάσμα μεταξύ υπολογιστικού μοντέλου και ανθρώπινης νευροφυσιολογίας: οι επαγγελματίες υγείας μπορούν να εξετάσουν τα κορυφαία χαρακτηριστικά και να αναγνωρίσουν οικείες εγκεφαλικές περιοχές αντί για αδιαφανή κανάλια ΕΕG. Παράλληλα, αξιοποιείται η πλούσια βιβλιογραφία fMRI και μελετών βλαβών για την ερμηνεία της επίδοσης BCI· λ.χ. εάν μια περιοχή που υποδεικνύει το SHAP είναι δυσλειτουργική σε έναν ασθενή (λόγω εγκεφαλικού ή νόσου), μπορεί να προβλεφθούν δυσχέρειες στην αντίστοιχη κλάση νοερής κίνησης ή να αναπροσαρμοστεί η εκπαίδευση με εναλλακτικές στρατηγικές.

Η επιβεβαίωση ότι οι αποφάσεις ενός BCI εντοπίζονται σε αναμενόμενες περιοχές μπορεί επίσης να καθοδηγήσει παρεμβάσεις νευροανάδρασης. Εάν γνωρίζουμε ότι το μοντέλο «βλέπει» κυρίως τον δεξιό paracentral lobule για τη νοερή κίνηση ποδιού, οι θεραπευτές μπορούν να σχεδιάσουν ασκήσεις που ενισχύουν ειδικά την ενεργοποίηση αυτής της ζώνης, βελτιώνοντας έτσι το σήμα που χρησιμοποιεί το BCI. Επιπλέον, προσφέρεται εργαλείο παρακολούθησης: στη νευροαποκατάσταση θα μπορούσε να εξετάζεται διαχρονικά αν η νοερή κίνηση του ασθενούς «ομαλοποιείται» — δηλαδή, αν τα σημαντικά χαρακτηριστικά του μοντέλου μετατοπίζονται προς τις κλασικές κινητικές περιοχές όσο προχωρά η εκπαίδευση. Συνολικά, η εναρμόνιση των εξηγήσεων βαθιάς μάθησης με μοτίβα ενεργοποίησης επιβεβαιωμένα από fMRI ενισχύει την κλινική ερμηνευσιμότητα και αξιοπιστία των συστημάτων BCI. Σε επιστημονικό επίπεδο, τα αποτελέσματα του μοντέλου παρέχουν ευκαιρία διερεύνησης λιγότερο κατανοητών όψεων της νοερής κίνησης. Λ.γ. η ανάδειξη του αριστερού κάτω μετωπιαίου λοβού (περιοχή Broca) στη νοερή κίνηση γλώσσας εγείρει ερωτήματα για τον ρόλο των συνδέσεων λόγου-κίνησης σε μη λεκτικές κινητικές φαντασίες. Το εύρημα αυτό-σε συμφωνία με προγενέστερες εργασίες [205]-υπονοεί ότι ακόμα και σε «καθαρές» κινητικές εργασίες, όπως η κίνηση γλώσσας, ο εγκέφαλος ενδέχεται να επιστρατεύει ανώτερες περιοχές, πιθανώς λόγω του διπλού ρόλου της γλώσσας στην ομιλία. Δεδομένου ότι η ανάλυσή μας ήταν data-driven, παρέχει ανεξάρτητη επιβεβαίωση πως τέτοια διαλειτουργική στρατολόγηση συμβαίνει σε υγιείς συμμετέχοντες εντός BCI. Γενικά, η ερμηνεύσιμη TN σε EEG μπορεί να αποκαλύψει δευτερεύουσες ενεργοποιήσεις που ίσως διαφεύγουν σε ανάλυση αισθητήρων αλλά γίνονται εμφανείς όταν συγκεντρώνονται κατά πηγή. Επομένως, πέρα από την επιβεβαίωση γνωστών προτύπων, η προσέγγισή μας μπορεί να γεννήσει υποθέσεις για πρόσθετα δίκτυα που εμπλέκονται στη νοερή κίνηση (π.χ. βρεγματικές ή προμετωπιαίες περιοχές που το μοντέλο θεώρησε σημαντικές, και θα μπορούσαν να διερευνηθούν περαιτέρω με νευροαπεικονιστικές μεθόδους).

#### Σύγκριση με την ερμηνεία ΕΕG σε επίπεδο αισθητήρων

Παραδοσιακά, η νοερή κίνηση στο EEG αναλύεται στο επίπεδο καναλιών (π.χ. μεταβολές ισχύος στα C3, C4, Cz κ.λπ.). Αν και αυτά τα τοπογραφικά πρότυπα έχουν περιγραφεί επαρκώς—με ERD/ERS πάνω από τα ετερόπλευρα κεντρικά ηλεκτρόδια για MI χεριού στερούνται ακριβούς νευροανατομικής εξειδίκευσης [201]. Τα σήματα αισθητήρων αποτελούν μίγμα πολλών φλοιϊκών πηγών λόγω αγωγής όγκου· ένα κανάλι EEG μπορεί να «συλλαμβάνει» δραστηριότητα από αρκετές γειτονικές περιοχές. Αυτό δημιουργεί ασάφεια: π.χ. μείωση ισχύος στο C3 αποδίδεται συνήθως σε ενεργοποίηση του αριστερού κινητικού φλοιού, αλλά θα μπορούσε να επηρεάζεται και από προκινητική ή βρεγματική δραστηριότητα που προβάλλεται στο ίδιο ηλεκτρόδιο.

Στη μελέτη μας, χρησιμοποιώντας χαρακτηριστικά πηγαιοτοποθετημένου EEG και την παρασημασιολογική διαίρεση Desikan–Killiany, άρθηκαν πολλές από αυτές τις αμφισημίες. Οι χάρτες ερμηνευσιμότητας επισημαίνουν ανατομικές περιοχές (όχι αισθητήρες), επιτρέποντας απευθείας αντιστοίχιση με γνωστές εγκεφαλικές δομές—σημαντική πρόοδος για την αναγνωσιμότητα. Δείξαμε ότι ό,τι παραδοσιακά εικάζονταν (π.χ. «ERD στο C4 σημαίνει ενεργοποίηση δεξιού κινητικού φλοιού κατά MI αριστερού χεριού») τώρα τεκμηριώνεται ρητά: το μοντέλο εστιάζει στο ROI του δεξιού προκεντρικού έλικα στα δοκιμαστικά του αριστερού χεριού, επιβεβαιώνοντας την πηγή του προτύπου EEG [197], [201]. Έτσι, δεν απλώς επικυρώνεται η κλασική ερμηνεία· βελτιώνεται—διακρίνουμε, π.χ., μεταξύ πρωτοταγούς κινητικής και προκινητικής συμβολής στην ίδια πλευρά, κάτι αδύνατο μόνο με αισθητήρες.

Επιπλέον, η προσέγγιση επιπέδου πηγής επέτρεψε την ανίχνευση συνεισφορών από περιοχές που μπορεί να διαφύγουν στην ανάλυση αισθητήρων. Χαρακτηριστικό παράδειγμα είναι η εμπλοκή της περιοχής Broca στη MI γλώσσας: κανένα τυπικό ηλεκτρόδιο δεν επικαλύπτει άμεσα τη Broca (μετωπιαία, πλάγια). Το αισθητηριακό EEG ίσως καταγράψει αμυδρά τη δραστηριότητά της σε πολλαπλά μετωπιαία κανάλια, καθιστώντας δύσκολη την ταυτοποίηση χωρίς ανακατασκευή πηγής. Η μέθοδός μας, όμως, απομόνωσε το σήμα και του απέδωσε σημαντικότητα. Ανάλογα, η διάκριση πολύ γειτονικών περιοχών (π.χ. πρόσθια vs. οπίσθια τράπεζα κεντρικής αύλακας ή SMA έναντι γειτονικών φλοιού) διευκολύνεται από τη συγκεκριμένη παρσελοποίηση. Η παραδοσιακή ανάλυση ενδέχεται να χαρακτηρίσει αδιακρίτως όλα τα κεντρικά ευρήματα ως «mu-ρυθμός στον αισθητικοκινητικό φλοιό», ενώ η δική μας δείχνει ποιο τμήμα (π.χ. περιοχή χεριού vs. ποδιού) κινεί το αποτέλεσμα. Πρόκειται για ουσιαστικό βήμα στην ερμηνευσιμότητα EEG—φέρνοντας τη χωρική ακρίβεια πιο κοντά στο fMRI, διατηρώντας την υψηλή χρονική διακριτική ικανότητα και την πρακτικότητα του EEG.

Σε σχέση με συμβατικές τεχνικές αποκωδικοποίησης EEG (όπως Common Spatial Patterns ή χειροποίητα χαρακτηριστικά καναλιών), ο συνδυασμός βαθιάς μάθησης + SHAP παρέχει λεπτομερέστερη εικόνα των νευροανατομικών θεμελίων. Οι κλασικές μέθοδοι BCI συχνά αποδίδουν απλώς μια κατηγορία ή, στην καλύτερη περίπτωση, έναν τοπογραφικό χάρτη βαρών που απαιτεί περαιτέρω ερμηνεία. Εδώ, πετύχαμε χαρτογράφηση από το εκπαιδευμένο μοντέλο πίσω σε εγκεφαλικές περιοχές, ώστε να συζητηθεί άμεσα σε όρους εγκεφαλικής λειτουργίας. Η σύζευξη μηχανικής μάθησης και νευροεπιστημονικής βιβλιογραφίας είναι πολύτιμη: διασφαλίζει ότι οι αποφάσεις του μοντέλου επικοινωνούνται στους νευροεπιστήμονες στη γλώσσα των φλοιϊκών περιοχών και δικτύων και επιτρέπει διασταύρωση με προηγούμενες μελέτες—κάτι σπάνιο σε «μαύρα κουτιά». Όταν οι εξηγήσεις του μοντέλου αποκλίνουν από τα αναμενόμενα πρότυπα, τέτοιες διαφορές μπορούν να υποδείζουν είτε ενδιαφέρουσα νευροδυναμική είτε πιθανά σφάλματα μοντέλου προσφέροντας, σε κάθε περίπτωση, χρήσιμη γνώση.

#### Περιορισμοί και επιφυλάξεις

Παρά την επιτυχή επαλήθευση της σημασίας των περιοχών με βάση τη γνωστή φυσιολογία, πρέπει να αναγνωριστούν αρκετοί περιορισμοί. Η ακρίβεια της εντοπισμού πηγών είναι

πεπερασμένη. Χρησιμοποιήσαμε τον τυπικό άτλαντα **Desikan–Killiany** και γενικό μοντέλο κεφαλής<sup>.</sup> σφάλματα στην ανακατασκευή μπορεί να αποδώσουν δραστηριότητα σε γειτονικές περιοχές. Π.χ. ο «χειρομορφικός θόλος» του precentral γειτνιάζει με τη ζώνη προσώπου υψηλή δραστηριότητα στη μία μπορεί να «διαρρεύσει» στο ROI της άλλης. Επιλέξαμε σχετικά λεπτή τμηματοποίηση (68 περιοχές), όμως η χωρική και πλέον και η χρονική ανάλυση παραμένει καλύτερη σε σχέση με το fMRI. Ως εκ τούτου, τα αποτελέσματα ερμηνευσιμότητας πρέπει να ιδωθούν ως ανάδειξη γενικών περιοχών, όχι ακριβούς θέσης σε σχέση με κάθε γύρης/αύλακος. Στην πράξη, η συμφωνία σε επίπεδο λοβού ή έλικος ήταν άριστη, χωρίς όμως να αποκλείεται κάποια ασάφεια (π.χ. τμήμα του superior parietal να αποδοθεί αντί για γειτονικό postcentral). Επιβεβαίωση σε λεπτότερη κλίμακα θα απαιτούσε fMRI ή ηλεκτροκορτικογραφία.

Χρονική γενίκευση τιμών SHAP. Εφαρμόσαμε τόσο γενικό μέσο όρο όσο και συγκεκριμένο «ενεργό παράθυρο» χρόνου· αυτό παρείχε ενδείξεις για το πότε είναι σημαντική μια περιοχή. Ωστόσο, τα EEG σήματα είναι ιδιαίτερα δυναμικά: διαφορετικά δίκτυα ενδέχεται να εμπλέκονται παροδικά σε διάφορες φάσεις του δοκιμίου (π.χ. πρώιμη οπτική επεξεργασία στο ινιακό, μεσαία κινητική προσομοίωση, όψιμη αυτοπαρακολούθηση μετωποβρεγματικά). Η ανάλυσή μας φωτίζει κυρίως το διαρκές τμήμα της νοερής κίνησης και ίσως υποτιμά ρόλους προετοιμασίας ή τερματισμού. Μελλοντικές εργασίες θα μπορούσαν να υιοθετήσουν λεπτότερη χρονικά ερμηνευσιμότητα (κυλιόμενα παράθυρα ή συμβάντο-σχετιζόμενη SHAP) για να καταγράψουν αυτά τα ακολουθιακά φαινόμενα. Στο δικό μας σύνολο, η στρατηγική «ενεργού παραθύρου» κάλυψε μερικώς το ζήτημα, ενισχύοντας κυρίως τις ίδιες περιοχές με την παγκόσμια ανάλυση.

#### Υποθέσεις του SHAP και πολυσυγγραμμικότητα.

Το SHAP (ιδίως το Kernel SHAP που χρησιμοποιήθηκε) προϋποθέτει ορισμένες απλοποιήσεις. Τα χαρακτηριστικά EEG (68 περιοχές × χρόνο) μπορεί να είναι ιδιαίτερα συσχετισμένα, επηρεάζοντας τη μοναδικότητα των αποδόσεων. Π.χ. τα σήματα δεξιού και αριστερού *M1* δεν είναι ανεξάρτητα—κοινά ταλαντωτικά ερεθίσματα ή αγωγή όγκου τα συσχετίζουν. Το SHAP επιδιώκει δίκαιη κατανομή βαρύτητας, παραμένει όμως πιθανότητα η πολυσυγγραμμικότητα να επηρεάζει ποια χαρακτηριστικά εμφανίζονται «κορυφαία». Πρακτικά, παρατηρήσαμε ότι το μοντέλο συχνά αποδίδει σημασία και στα δύο μέλη ενός ετερόπλευρου ζεύγους, με αντίθετα πρόσημα· φαινόμενο σύμφωνο με τη «τραμπάλα» ισχύος στα ΜΙ. Οι αποδόσεις SHAP φαίνονται εύλογες, χρειάζεται όμως προσοχή στην υπερερμηνεία της ακριβούς σειράς κατάταξης δευτερευουσών περιοχών. Αν, λ.χ., ο αριστερός premotor και ο αριστερός primary motor είναι και οι δύο ενεργοί για ΜΙ δεξιού χεριού, το SHAP μπορεί να μοιράσει τη βαρύτητα κάπως αυθαίρετα, εκτός αν το μοντέλο διακρίνει σαφώς τις συνεισφορές τους. Επομένως, ένας κατάλογος «top-5» περιοχών πρέπει να εξετάζεται παράλληλα με το γνωστό δίκτυο περιοχών που τείνουν να συν-ενεργοποιούνται.

Ένας πρόσθετος περιορισμός είναι ότι η ανάλυσή μας περιορίστηκε αποκλειστικά στον φλοιό, μολονότι υποφλοιώδεις δομές—όπως η παρεγκεφαλίδα και τα βασικά γάγγλια συμμετέχουν τόσο στη νοερή όσο και στην πραγματική κίνηση. Η παρεγκεφαλίδα, ειδικότερα, εμφανίζει συνήθως **ομόπλευρη** ενεργοποίηση κατά την εκτέλεση κινήσεων άκρων και έχει αναφερθεί ότι εμπλέκεται και στη νοερή κίνηση [199], [202]. Ωστόσο, το επιφανειακό ΕΕG είναι σχετικά αναίσθητο σε βαθιές πηγές και το μοντέλο πηγών μας δεν περιλάμβανε παρεγκεφαλιδικά πεδία· έτσι, τα αποτελέσματα ερμηνευσιμότητας δεν θα μπορούσαν να αποτυπώσουν συμμετοχή της παρεγκεφαλίδας. Αυτό ίσως εξηγεί γιατί οι αρνητικές αποδόσεις του μοντέλου για MI χεριού δεν περιλάμβαναν ρητά «ετερόπλευρη παρεγκεφαλίδα», παρότι θα μπορούσε να αναμένεται (λ.χ. η νοερή κίνηση αριστερού χεριού εμπλέκει συχνά δεξιά παρεγκεφαλίδα). Στο EEG, τα παρεγκεφαλιδικά σήματα είναι πολύ εξασθενημένα ώστε να ανιχνευθούν με αξιοπιστία, οπότε το μοντέλο τα αγνόησε φυσιολογικά. Πρόκειται για περιορισμό ως προς την πλήρη αποτύπωση του δικτύου νοερής κίνησης—όχι όμως σφάλμα του μοντέλου, αλλά περιορισμό της ίδιας της μεθόδου καταγραφής. Μελλοντικές εργασίες θα μπορούσαν να ενσωματώσουν πολυτροπικές μεθόδους (π.χ. συγχώνευση EEG–fMRI ή EEG με μαγνητόμετρα) ώστε να συμπεριλάβουν αυτές τις υποφλοιώδεις συνεισφορές. Προς το παρόν, είναι καθησυχαστικό ότι τουλάχιστον η φλοιϊκή συνιστώσα του δικτύου αποδόθηκε με πιστότητα.

#### Ευθυγράμμιση βαθιάς μάθησης και νευροεπιστήμης – μελλοντικές κατευθύνσεις

Τα ευρήματά μας δείχνουν πώς η **ερμηνεύσιμη βαθιά μάθηση** μπορεί να λειτουργήσει τόσο ως εργαλείο επαλήθευσης όσο και ανακάλυψης στην BCI έρευνα που καθοδηγείται από τη νευροεπιστήμη. Η σαφής αντιστοιχία ανάμεσα στις σημαντικές περιοχές που αντλεί το μοντέλο και στη βιβλιογραφία fMRI προσδίδει αξιοπιστία στη χρήση τέτοιων μοντέλων στην πράξη· αποδεικνύει ότι το «μαύρο κουτί» είναι πράγματι διαπερατό όταν ιδωθεί υπό το πρίσμα της γνωστής εγκεφαλικής λειτουργίας. Η σύζευξη αυτή αξίζει περαιτέρω αξιοποίηση.

- Ενσωμάτωση νευροβιολογικών priors. Ένας μελλοντικός άξονας είναι η ενσωμάτωση νευροεπιστημονικών προτέρων απευθείας στην εκπαίδευση: να δομηθεί το δίκτυο έτσι ώστε μονάδες ή στρώματα να αντιστοιχούν σε ανατομικές περιοχές, ή να επιβληθεί κανονικοποίηση που ευνοεί κατανομές βαρών σύμφωνες με γνωστά πρότυπα ενεργοποίησης. Αν το μοντέλο γνωρίζει εκ των προτέρων ότι η νοερή κίνηση αριστερού και δεξιού χεριού πρέπει να διαχωρίζεται βάσει πλευρίωσης, ίσως μάθει ταχύτερα ή πιο ανθεκτικά σε μικρά δείγματα. Παρ' όλα αυτά, τα αποτελέσματά μας δείχνουν ότι το μοντέλο δύναται να «ανακαλύψει» αυτά τα πρότυπα αυτοδύναμα, στοιχείο ενθαρρυντικό για data-driven προσέγγιση.
- Οξύτερος εντοπισμός πηγών. Για περαιτέρω βελτίωση της νευροανατομικής ερμηνευσιμότητας, θα μπορούσε να χρησιμοποιηθεί υψηλότερης πυκνότητας EEG ή εξατομικευμένα μοντέλα κεφαλής. Η μετάβαση από 22 κανάλια (όπως στο BCI IV 2a) σε 64 ή 128 θα επέτρεπε τη διάσπαση ορισμένων περιοχών DK σε λεπτότερα υποτμήματα, διακρίνοντας, π.χ., ραχιαίο έναντι κοιλιακού premotor φλοιού ή σωματοαισθητικό έναντι κινητικού τμήματος εντός των postcentral/precentral ελίκων. Θα μπορούσε να υιοθετηθεί λεπτομερέστερη παρσελοποίηση (όπως ο άτλαντας Destrieux ή λειτουργικές παρσελώσεις από resting-state fMRI). Η αύξηση χαρακτηριστικών απαιτεί διαχείριση μεγαλύτερης πολυπλοκότητας για το μοντέλο και τους υπολογισμούς SHAP, αλλά θα μπορούσε να αποδώσει ακόμα πιο λεπτομερείς χάρτες—π.χ. να απομονωθεί το «hand-knob» του precentral από το υπόλοιπο τμήμα, αποκαλύπτοντας ακόμη ακριβέστερο σωματοτοπικό SHAP-mapping.
- Ένας ακόμη πολλά υποσχόμενος άξονας είναι η επέκταση της ερμηνευσιμότητας ώστε να συλλαμβάνει αλληλεπιδράσεις μεταξύ περιοχών. Η παρούσα ανάλυση θεώρησε τη σημασία κάθε περιοχής ανεξάρτητη, όμως τα πρότυπα εγκεφαλικής δραστηριότητας συχνά βασίζονται σε συνεργασίες (π.χ. συν-ενεργοποίηση

προκινητικού και βρεγματικού φλοιού). Μέθοδοι που εξηγούν διμερείς ή υψηλότερης τάξης αλληλεπιδράσεις θα μπορούσαν, λόγου χάρη, να αποκαλύψουν ότι «η ταυτόχρονη ενεργοποίηση του αριστερού Μ1 και του οπίσθιου αριστερού βρεγματικού φλοιού αποτελεί ισχυρό δείκτη νοερής κίνησης δεξιού χεριού»—εύρημα συμβατό με μελέτες που τονίζουν το εμπρόσθιο-βρεγματικό δίκτυο στη νοερή κίνηση [33]. Η ανάδειξη τέτοιων αλληλεπιδράσεων θα ευθυγραμμίσει περαιτέρω τη λογική του μοντέλου με δικτυακές έννοιες (π.χ. ραχιαίο ρεύμα στον κινητικό σχεδιασμό). Τεχνικές όπως οι SHAP interaction values ή ένας rule-based learner επάνω στις αποδόσεις του μοντέλου μπορεί να εξάγουν αυτές τις γνώσεις.

- Ως προς την απόδοση BCI, ένα ενδιαφέρον μελλοντικό βήμα είναι να αξιοποιηθούν οι ερμηνεύσιμες γνώσεις σε επίπεδο περιοχών για βελτιωμένη ανατροφοδότηση προς τους χρήστες. Αντί να ειπωθεί «η ισχύς EEG στη ζώνη X στο ηλεκτρόδιο C3 είναι χαμηλή», θα μπορούσε να δοθεί η οδηγία «ενεργοποίησε περισσότερο τον δεξιό κινητικό φλοιό—προσπάθησε να "νιώσεις" το αριστερό σου χέρι να κινείται». Πρόκειται για πολύ πιο διαισθητική καθοδήγηση, θεμελιωμένη στην ίδια την εγκεφαλική ανατομία του ατόμου. Κλείνοντας τον βρόχο με ανατομικά πληροφορημένη ανατροφοδότηση, οι χρήστες ίσως ρυθμίζουν αποτελεσματικότερα τις ορθές περιοχές, ενισχύοντας τον έλεγχο BCI. Ορισμένα συστήματα νευροανάδρασης ήδη στοχεύουν φλοιϊκές περιοχές μέσω fMRI ή εκπαίδευσης πηγών EEG· η προσέγγισή μας παρέχει τρόπο εξαγωγής αυτών των «στόχων» απευθείας από τον ταξινομητή.
- Τέλος, απαιτείται έλεγχος γενίκευσης (generalization) των ευρημάτων σε ευρύτερους πληθυσμούς και συνθήκες. Η τρέχουσα επικύρωση αφορά υγιείς εθελοντές σε κλασικές εργασίες νοερής κίνησης. Σε κλινικούς πληθυσμούς (π.χ. ασθενείς με ημιπάρεση μετά από εγκεφαλικό) τα πρότυπα ενεργοποίησης ενδέχεται να μεταβληθούν—π.χ. ο ετερόπλευρος M1 να μην είναι διαθέσιμος και να αναλαμβάνουν ομόπλευρες ή δευτερεύουσες περιοχές (premotor, ή το ετερόπλευρο ημισφαίριο). Σε τέτοιες περιπτώσεις, οι αποδόσεις SHAP του μοντέλου μπορεί να αποκλίνουν από τα «εγχειριδιακά» πρότυπα. Η εφαρμογή της ίδιας ανάλυσης σε δεδομένα ασθενών θα έδειχνε αν τα σημαντικά χαρακτηριστικά του μοντέλου αντιστοιχούν σε αντιρροπιστικές νευροπλαστικές αλλαγές που είναι γνωστές από κλινικές μελέτες. Η εναρμόνιση της ερμηνευσιμότητας του μοντέλου με τα fMRI ή τα δεδομένα βλαβών κάθε ασθενούς θα μπορούσε να εξατομικεύσει τη θεραπεία BCI: αν ο ασθενής χρησιμοποιεί μη συμβατική περιοχή για να φαντάζεται κινήσεις, το μοντέλο θα αναπροσαρμοστεί ή ο ασθενής θα επανεκπαιδευτεί ανάλογα.

## 6. Συζήτηση και Επίλογος

Η παρούσα διατριβή διερεύνησε τη δυνατότητα βελτίωσης της κατανόησης και ερμηνευσιμότητας των μοντέλων Τεχνητής Νοημοσύνης σε εφαρμογές ηλεκτροεγκεφαλογραφίας (ΗΕΓ), εστιάζοντας ιδίως στις διεπαφές εγκεφάλου-υπολογιστή (BCI). Το έργο αυτό αναπτύχθηκε σε τρεις κύριους άξονες: την εφαρμογή ανάλυσης χρόνουσυχνότητας (κεφάλαιο 3), την ανασκόπηση και ανάλυση των μεθόδων εντοπισμού πηγών ΗΕΓ (κεφάλαιο 4) και την εφαρμογή των παραπάνω, σε συνδυασμό με ερμηνεύσιμη βαθιά μάθηση, για την ανάπτυξη ενός εξηγήσιμου συστήματος BCI νοερής κίνησης (κεφάλαιο 5).

## 6.1 Συνοπτικά Ευρήματα και Συνεισφορές

Στο πρώτο μέρος της εργασίας (Κεφάλαιο 3), επιδείχθηκε η αποτελεσματικότητα της ανάλυσης χρόνου-συχνότητας σε συνδυασμό με ένα υπολογιστικά αποδοτικό συνελικτικό νευρωνικό δίκτυο για την ταξινόμηση σημάτων ΗΕΓ σε ένα διαφορετικό πλαίσιο: την ανίχνευση νοητικής κόπωσης. Αυτό το τμήμα τόνισε την αξία της χρήσης "μηχανικής" χαρακτηριστικών (feature engineering) μέσω της ανάλυσης TFR και ανέδειξε πώς ακόμη και ρηχές αρχιτεκτονικές, σε συνδυασμό με κατάλληλα προεπεξεργασμένα δεδομένα και μεθόδους ερμηνευσιμότητας όπως το SHAP, μπορούν να επιτύχουν υψηλή απόδοση (97% ακρίβεια) με χαμηλό υπολογιστικό κόστος, εντοπίζοντας παράλληλα φασματικά πρότυπα (στις ζώνες θήτα/άλφα) συμβατά με τη βιβλιογραφία της κόπωσης. Η επιτυχία αυτή, ακόμη και σε ένα υποσύνολο δεδομένων (22 υποκείμενα) και με χρήση μόνο CPU, υπογράμμισε τη δυνατότητα ανάπτυξης συστημάτων TN για ΗΕΓ με δυνατότητες πραγματικού χρόνου και εφαρμογές σε συσκευές EDGE-AI.

Το δεύτερο μέρος (Κεφάλαιο 4) παρείχε μια κριτική επισκόπηση των τρεχουσών μεθόδων εντοπισμού πηγών ΗΕΓ. Αναλύθηκαν οι θεωρητικές αρχές, οι διαφορετικές κατηγορίες αλγορίθμων (π.χ. beamformers, αναλύσεις ελάχιστης νόρμας), οι πρακτικές εφαρμογές τους στη νευροεπιστήμη και τις κλινικές μελέτες, καθώς και οι εγγενείς περιορισμοί τους (π.χ. πρόβλημα προς τα εμπρός και προς τα πίσω, ανάγκη για ακριβή μοντέλα κεφαλής, μειωμένη ευαισθησία σε βαθιές πηγές). Αυτή η ανασκόπηση θεμελίωσε τη σημασία της μετάβασης από τον χώρο των αισθητήρων ΗΕΓ στον χώρο των εγκεφαλικών πηγών για την επίτευξη ανατομικά έγκυρης ερμηνείας και δικαιολόγησε την επιλογή του beamformer source localization ως μια κατάλληλη προσέγγιση για την εκτίμηση φλοιϊκής δραστηριότητας στις επόμενες ενότητες.

Το τρίτο και κεντρικό μέρος της διατριβής (Κεφάλαιο 5) εφάρμοσε τα παραπάνω σε ένα ολοκληρωμένο πλαίσιο για την ταξινόμηση σημάτων ΗΕΓ κατά τη νοερή κίνηση, με στόχο την ενίσχυση της ερμηνευσιμότητας των συστημάτων BCI. Χρησιμοποιόντας τον εντοπισμό πηγής με beamformer σε περιοχές του άτλαντα Desikan–Killiany και συνελικτικά νευρωνικά δίκτυα, επιτεύχθηκε ακρίβεια ανταγωνιστική σε σχέση με τα κλασικά μοντέλα BCI. Το ουσιαστικό επίτευγμα, ωστόσο, ήταν η ενσωμάτωση ερμηνευτικών τεχνικών όπως οι SHAP και Integrated Gradients. Αυτές οι τεχνικές δεν περιορίστηκαν στην ανάδειξη σημαντικών καναλιών ή χρονικών σημείων, αλλά, χάρη στην προηγηθείσα ανάλυση source localization, αποκάλυψαν ένα λειτουργικό δίκτυο ενεργών φλοιϊκών περιοχών.

Η εφαρμογή στο BCI νοερής κίνησης λειτούργησε ως άριστο πεδίο επίδειξης του προτεινόμενου πλαισίου, καθώς η νευροφυσιολογία της κινητικής φαντασίας είναι σχετικά καλά τεκμηριωμένη. Η ανάλυση ερμηνευσιμότητας (ιδίως SHAP) ανέδειξε ένα σαφές σωματοτοπικό πρότυπο, σύμφωνο με τη γνωστή οργάνωση του κινητικού φλοιού και τα ευρήματα από μελέτες fMRI. Συγκεκριμένα, η νοερή κίνηση χεριών συνδέθηκε κυρίως με ετερόπλευρες προκεντρικές περιοχές, η κίνηση ποδιών με διμερείς παρακεντρικούς λοβούς και η κίνηση γλώσσας με κατώτερες προκεντρικές ζώνες (συμπεριλαμβανομένης της περιοχής Broca). Αυτή η ισχυρή συμφωνία με την εδραιωμένη νευροεπιστημονική γνώση είναι κρίσιμη: επιβεβαιώνει ότι ο ταξινομητής βασίζεται σε γνήσια νευροανατομικά σήματα και όχι σε τυχαίους θορύβους ή τεχνητά μοτίβα ηλεκτροδίων.

Επιπλέον, το μοντέλο κατέγραψε με επιτυχία τα φαινόμενα πλευρίωσης, δείχνοντας ετερόπλευρη κυριαρχία για τη νοερή κίνηση χεριών, συνάδοντας με τον διαημισφαιρικό έλεγχο των άκρων. Η δυνατότητα σύνδεσης αυτών των παρατηρήσεων επιπέδου αισθητήρων (ερμηνευόμενων πλέον σε επίπεδο πηγών) με συγκεκριμένες φλοιϊκές περιοχές ενισχύει την πεποίθηση ότι οι εσωτερικές αναπαραστάσεις του μοντέλου αντανακλούν την αληθή φλοιϊκή φυσιολογία, συμπεριλαμβανομένων λεπτών διαφοροποιήσεων που σχετίζονται με την προτίμηση χειρός.

Σημαντικό στοιχείο της συνεισφοράς μας είναι η άρση της ασάφειας της παραδοσιακής ανάλυσης ΗΕΓ σε επίπεδο αισθητήρων. Ενώ τα τοπογραφικά πρότυπα (π.χ. ERD/ERS σε κεντρικά ηλεκτρόδια) είναι γνωστά, η πηγαιοτοποθετημένη προσέγγιση επιτρέπει την άμεση αντιστοίχιση της σημαντικότητας των χαρακτηριστικών με ανατομικές περιοχές (ROIs). Αυτό βελτιώνει την αναγνωσιμότητα και επιτρέπει διακρίσεις μεταξύ γειτονικών ή λειτουργικά διαφορετικών περιοχών (π.χ. πρωτοταγής κινητικός έναντι προκινητικού φλοιού) που είναι δύσκολες με ανάλυση μόνο αισθητήρων. Η ανάδειξη περιοχών όπως η Broca στην κίνηση γλώσσας, που δεν καλύπτεται άμεσα από τυπικά ηλεκτρόδια, καταδεικνύει τη δυνατότητα της εντοπισμού πηγών σε συνδυασμό με την ερμηνευσιμότητα να αποκαλύπτουν συνεισφορές που διαφεύγουν σε επίπεδο αισθητήρων.

## 6.2 Σημασία για Εφαρμογές και Νευροεπιστήμη

Η σύγκλιση των ερμηνεύσιμων εξόδων του μοντέλου μας με την εδραιωμένη νευροεπιστημονική γνώση έχει σημαντικές κλινικές και επιστημονικές επιπτώσεις. Για κλινικές εφαρμογές BCI (π.χ. νευροαποκατάσταση), η δυνατότητα να επαληθεύεται ότι το μοντέλο βασίζει τις αποφάσεις του στις αναμενόμενες, φυσιολογικώς σχετικές περιοχές (π.χ. τον κινητικό φλοιό) ενισχύει την εμπιστοσύνη στη χρήση του. Επιτρέπει στους κλινικούς να κατανοήσουν πώς λειτουργεί το BCI και να παρέχουν ανατομικά πληροφορημένη νευροανάδραση στους ασθενείς (π.χ. "ενεργοποίησε την περιοχή Χ") αντί για ασαφείς οδηγίες. Παράλληλα, οι χάρτες ερμηνευσιμότητας μπορούν να χρησιμοποιηθούν για την παρακολούθηση νευροπλαστικών αλλαγών, εντοπίζοντας μετατοπίσεις της ενεργοποίησης κατά τη διάρκεια της αποκατάστασης. Γεφυρώνεται έτσι το χάσμα μεταξύ της υψηλής απόδοσης της TN και της ανάγκης των κλινικών για διαφάνεια και φυσιολογική εγκυρότητα.

Σε επιστημονικό επίπεδο, το πλαίσιο προσφέρει ένα ισχυρό εργαλείο για τη διερεύνηση των νευρωνικών μηχανισμών. Πέρα από την επιβεβαίωση γνωστών προτύπων, η data-driven ανάλυση μπορεί να αποκαλύψει λιγότερο κατανοητές συνεισφορές (όπως η εμπλοκή της Broca στη νοερή κίνηση γλώσσας) ή να γεννήσει υποθέσεις για τη συμμετοχή άλλων περιοχών (π.χ. βρεγματικών ή προμετωπιαίων) που το μοντέλο θεωρεί σημαντικές, ωθώντας περαιτέρω νευροαπεικονιστικές έρευνες. Η ευθυγράμμιση των εξηγήσεων του μοντέλου με όρους εγκεφαλικών περιοχών διευκολύνει τον διάλογο μεταξύ ερευνητών TN και νευροεπιστημόνων και επιτρέπει τη συσχέτιση των ευρημάτων BCI με την πλούσια βιβλιογραφία από fMRI και μελέτες βλαβών.

## 6.3 Περιορισμοί και Επιφυλάζεις

Παρά τις παραπάνω επιτυχίες, είναι σημαντικό να αναγνωριστούν οι περιορισμοί των εφαρμοζόμενων μεθοδολογιών. Πρώτον, η ακρίβεια της source localization ΗΕΓ είναι πεπερασμένη. Η χρήση τυπικών ατλάντων (όπως ο Desikan–Killiany) και γενικών μοντέλων κεφαλής εισάγει εγγενείς ανακρίβειες. Ενώ η συμφωνία σε επίπεδο λοβού ή μεγάλης έλικας ήταν εξαιρετική, η λεπτή διάκριση μεταξύ πολύ γειτονικών δομών ή η ακριβής τοποθέτηση εντός μιας έλικας παραμένει πρόκληση. Η χωρική ανάλυση, αν και βελτιωμένη σε σχέση με την ανάλυση αισθητήρων, δεν φτάνει αυτή του fMRI. Δεύτερον, όπως επισημάνθηκε και στην ανάλυση κόπωσης, το μέγεθος των συνόλων δεδομένων (ιδίως για το BCI όπου χρησιμοποιήθηκε τυπικό dataset όπως το BCI IV 2a) μπορεί να θεωρηθεί μέτριο για εφαρμογές βαθιάς μάθησης, αν και συγκρίσιμο με άλλες μελέτες BCI. Η γενίκευση σε μεγαλύτερους και πιο ετερογενείς πληθυσμούς απαιτεί περαιτέρω έρευνα. Τρίτον, η εγγενής αδυναμία του επιφανειακού ΗΕΓ να συλλάβει με αξιοπιστία σήματα από βαθιές εγκεφαλικές δομές (παρεγκεφαλίδα, βασικά γάγγλια) σημαίνει ότι η ανάλυσή μας περιορίστηκε στον φλοιό, παραβλέποντας τη συμβολή αυτών των περιοχών στο δίκτυο νοερής κίνησης. Τέταρτον, η χρονική ανάλυση εστιάστηκε σε ένα κύριο παράθυρο δραστηριότητας, πιθανώς αγνοώντας δυναμικά φαινόμενα και διακριτές φάσεις εντός του δοκιμίου (π.χ. προετοιμασία, τερματισμός) που ενδέχεται να εμπλέκουν διαφορετικά δίκτυα. Τέλος, οι ίδιες οι τεχνικές ερμηνευσιμότητας (όπως το Kernel SHAP) έχουν υποθέσεις, ιδίως όσον αφορά την ανεξαρτησία των χαρακτηριστικών. Η πιθανή πολυσυγγραμμικότητα στα χαρακτηριστικά πηγών μπορεί να επηρεάσει την ακριβή κατανομή της σημασίας, αν και τα κύρια ευρήματά μας παρέμειναν βιολογικά εύλογα.

## 6.4 Προοπτικές Μελλοντικής Έρευνας

Οι παραπάνω περιορισμοί και ταυτόχρονα οι επιτυχίες του προτεινόμενου πλαισίου ανοίγουν πλούσιες κατευθύνσεις για μελλοντική έρευνα. Μία άμεση κατεύθυνση αφορά τη βελτίωση της χωρικής και χρονικής ανάλυσης του ΗΕΓ. Η χρήση συστημάτων υψηλότερης πυκνότητας (με 128 ή περισσότερα κανάλια), σε συνδυασμό με εξατομικευμένα μοντέλα κεφαλής και λεπτομερέστερους ανατομικούς ή λειτουργικούς ατλάντες (όπως ο Destrieux ή λειτουργικές παρσελώσεις από resting-state fMRI), μπορεί να αυζήσει σημαντικά την ακρίβεια της εντοπισμού πηγών και, κατ' επέκταση, την ανατομική εξειδίκευση των ερμηνεύσιμων χαρακτηριστικών. Παράλληλα, η ανάπτυξη χρονικά λεπτομερέστερων μεθόδων ερμηνευσιμότητας θα μπορούσε να αποτυπώσει τη δυναμική των εγκεφαλικών δικτύων καθ' όλη τη διάρκεια ενός δοκιμίου, φωτίζοντας διακριτές φάσεις όπως η προετοιμασία ή ο τερματισμός μιας νοερής κίνησης, οι οποίες ενδέχεται να εμπλέκουν διαφορετικά δίκτυα από την κύρια εκτέλεση. Συναφώς με την πληρέστερη απεικόνιση της εγκεφαλικής λειτουργίας, η ενσωμάτωση πολυτροπικών δεδομένων αποτελεί σημαντική προοπτική. Ο συνδυασμός του ΗΕΓ με άλλες νευροαπεικονιστικές μεθόδους, όπως το fMRI ή το MEG, ή η χρήση

προηγμένων μοντέλων εντοπισμού πηγών που περιλαμβάνουν υποφλοιώδεις δομές, μπορεί να δώσει μια πιο ολοκληρωμένη εικόνα των εμπλεκόμενων δικτύων, συμπεριλαμβάνοντας περιοχές όπως η παρεγκεφαλίδα και τα βασικά γάγγλια, τις οποίες το επιφανειακό ΗΕΓ δεν συλλαμβάνει επαρκώς. Πέρα από τη βελτίωση της ποιότητας των δεδομένων και του εντοπισμού πηγών, η διερεύνηση προηγμένων τεχνικών ερμηνευσιμότητας μπορεί να προσφέρει βαθύτερες γνώσεις. Για παράδειγμα, η διερεύνηση αλληλεπιδράσεων μεταξύ εγκεφαλικών περιοχών μέσω SHAP interaction values ή άλλων μεθόδων ανάλυσης δικτύων θα επέτρεπε την αποκάλυψη της συνεργασίας μεταξύ περιοχών, προχωρώντας πέρα από την απλή σπουδαιότητα μεμονωμένων ζωνών και ευθυγραμμίζοντας την ερμηνεία με δικτυακές έννοιες της εγκεφαλικής λειτουργίας. Μια άλλη σημαντική κατεύθυνση συνίσταται στην άμεση ενσωμάτωση νευροεπιστημονικών προτέρων (priors) στην αργιτεκτονική ή τη διαδικασία εκπαίδευσης των νευρωνικών δικτύων. Η ενσωμάτωση γνώσης για την εγκεφαλική λειτουργία (π.χ. γνωστά πρότυπα συνδεσιμότητας ή πλευρίωσης) μπορεί δυνητικά να βελτιώσει την απόδοση, την ανθεκτικότητα σε μικρά δείγματα και την εγγενή ερμηνευσιμότητα του μοντέλου, επιτρέποντας την πιο αποτελεσματική εκμετάλλευση των διαθέσιμων δεδομένων. Προφανής κλινική προοπτική αποτελεί η εφαρμογή του πλαισίου σε κλινικούς πληθυσμούς. Η διερεύνηση ασθενών (π.χ. με εγκεφαλικό, σκλήρυνση κατά πλάκας) θα επιτρέψει τη χαρτογράφηση των μεταβαλλόμενων ή αντιρροπιστικών νευρωνικών μηχανισμών που σχετίζονται με την πάθηση και, κατ' επέκταση, την εξατομίκευση των παρεμβάσεων BCI και νευροανάδρασης με βάση τα συγκεκριμένα πρότυπα ενεργοποίησης κάθε ασθενούς. Τέλος, το προτεινόμενο πλαίσιο, συνδυάζοντας ανάλυση χρόνου-συχνότητας, εντοπισμό πηγών, βαθιά μάθηση και ερμηνευσιμότητα, μπορεί να επεκταθεί σε ευρύ φάσμα άλλων εργασιών ταξινόμησης ΗΕΓ. Αυτό περιλαμβάνει την ανάλυση γνωστικών φορτίων, συναισθηματικών καταστάσεων, σταδίων ύπνου ή την ανίχνευση παθολογικών καταστάσεων όπως οι επιληπτικές κρίσεις, με στόχο την αποκάλυση των υποκείμενων νευρωνικών συσχετίσεων, όπως άλλωστε καταδείχθηκε με την ανάλυση κόπωσης στο Κεφάλαιο 3.

## 6.5 Επίλογος

Συνοψίζοντας, η παρούσα διατριβή παρουσίασε ένα συνεκτικό πλαίσιο που αξιοποιεί την ανάλυση χρόνου-συχνότητας, τον εντοπισμό πηγών ΗΕΓ, τη βαθιά μάθηση και τις ερμηνευτικές τεχνικές για την ανάπτυξη υψηλής απόδοσης μοντέλων ΗΕΓ που είναι ταυτόχρονα ανατομικά εξηγήσιμα. Μέσω της εφαρμογής σε ένα σύστημα BCI νοερής κίνησης, αποδείχθηκε ότι είναι εφικτό να γεφυρωθεί το παραδοσιακό δίλημμα μεταξύ ακρίβειας (black box) και ερμηνευσιμότητας. Η δυνατότητα να ανατρέχουμε στις αποφάσεις του μοντέλου και να τις αντιστοιχίζουμε με συγκεκριμένες εγκεφαλικές περιοχές, συμβατές με τη νευροεπιστημονική βιβλιογραφία, είναι ένα κρίσιμο βήμα προς την κλινική υιοθέτηση των συστημάτων TN σε εφαρμογές ΗΕΓ και προς την προώθηση της θεμελιώδους νευροεπιστημονικής γνώσης.

Η προτεινόμενη μεθοδολογία θεμελιώνει μια νέα γενιά διεπαφών εγκεφάλου-υπολογιστή και νευροεπιστημονικών εργαλείων, όπου η κατανόηση της υποκείμενης νευροφυσιολογίας βρίσκεται στον πυρήνα του σχεδιασμού και της χρήσης. Αυτό το όραμα οδηγεί σε συστήματα όπου η συνεργασία μεταξύ ανθρώπου και Τεχνητής Νοημοσύνης είναι αμφίδρομη και διαφανής, επιταχύνοντας τη μετάβαση των τεχνολογιών ΗΕΓ από το εργαστήριο στην κλινική πράξη και την καθημερινότητα, προς όφελος ασθενών και χρηστών.

## Γλωσσάρι Συντομογραφιών

- **Ag/AgCl:** Άργυρος/Χλωριούχος Άργυρος
- AUC / AUROC: Εμβαδόν κάτω από την καμπύλη ROC (Receiver Operating Characteristic)
- **BAE:** Bayesian Approximation Error (Σφάλμα Μπεϋζιανής Προσέγγισης)
- ΒΜ: Βαθιά Μάθηση
- CG-FEM: Συνεχής Μέθοδος Galerkin-FEM (Finite Element Method)
- CNN(s): Συνελικτικό Νευρωνικό Δίκτυο / Συνελικτικά Νευρωνικά Δίκτυα
- **ConvDip:** (Όνομα συγκεκριμένου Συνελικτικού Νευρωνικού Δικτύου)
- ΔΔΕΥ (BCI): Διεπαφές Εγκεφάλου-Υπολογιστή (αναφέρεται συχνά ως BCI στο κείμενο)
- DeepLIFT: (Deep Learning Important FeaTures Σημαντικά Χαρακτηριστικά Βαθιάς Μάθησης)
- DST-DAE: (Όνομα συγκεκριμένου Αυτόματου Κωδικοποιητή Αποθορυβοποίησης)
- dw-MNE: Σταθμισμένη κατά Βάθος Εκτίμηση Ελάχιστης Νόρμας
- ECDs: Ισοδύναμα Ρεύματα Διπόλων
- eLORETA: Ακριβής Ηλεκτρομαγνητική Τομογραφία Χαμηλής Ανάλυσης
- ENY (CSF): Εγκεφαλονωτιαίο Υγρό
- ΗΕΓ (EEG): Ηλεκτροεγκεφαλογράφημα / Ηλεκτροεγκεφαλογραφία
- ΗΚΓ (ECG/EKG): Ηλεκτροκαρδιογραφία
- ΗΜΓ (EMG): Ηλεκτρομυογραφία
- ESL: EEG Source Localization (Απεικόνιση Ηλεκτρικής Πηγής ΗΕΓ)
- **EOG:** Ηλεκτροοφθαλμογραφία
- ERPs: Δυναμικά Σχετιζόμενα με Γεγονότα
- FC: Πλήρως Συνδεδεμένο Επίπεδο
- FEM: Μέθοδος Πεπερασμένων Στοιχείων
- FIR: Φίλτρο Πεπερασμένης Κρουστικής Απόκρισης
- fMRI: Λειτουργική Απεικόνιση Μαγνητικού Συντονισμού
- Grad-CAM: (Gradient-weighted Class Activation Mapping)
- ICA: Ανάλυση Ανεξάρτητων Συνιστωσών
- IG: Ολοκληρωμένες Κλίσεις
- IIR: Φίλτρο Άπειρης Κρουστικής Απόκρισης
- Iz: (Σημείο Inion θέση ηλεκτροδίου ΗΕΓ)
- KA (MI): Κινητική Απεικόνιση
- LCMV: Γραμμικά Περιορισμένη Ελάχιστη Διακύμανση
- LD-ESI: Εντοπισμός Πηγής ΗΕΓ Χαμηλής Πυκνότητας
- LFM: Πίνακας Πεδίου Αγωγιμότητας (Leadfield Matrix)
- LIME: Local Interpretable Model-Agnostic Explanations
- LORETA: Ηλεκτρομαγνητική Τομογραφία Εγκεφάλου Χαμηλής Ανάλυσης
- LSTM: Μονάδες Μακράς Βραχυπρόθεσμης Μνήμης
- Μ1 (Κινητικός Φλοιός): Πρωτοταγής Κινητικός Φλοιός
- Μ1, Μ2 (Ηλεκτρόδια): Μαστοειδή Ηλεκτρόδια
- MEG: Μαγνητοεγκεφαλογραφία
- ΜΝΕ: Εκτίμηση Ελάχιστης Νόρμας
- MNI: Montreal Neurological Institute (Νευρολογικό Ινστιτούτο του Μόντρεαλ)
- MT (MRI): Μαγνητική Τομογραφία

- **PCANet:** (Δίκτυο Ανάλυσης Κύριων Συνιστωσών)
- ΡΜ: Προβολική Μέθοδος
- **PPV:** Θετική Προγνωστική Αξία
- **PSD:** Φασματική Πυκνότητα Ισχύος
- **ROIs:** Περιοχές Ενδιαφέροντος (Regions of Interest)
- **ROC:** Καμπύλη Λειτουργικού Χαρακτηριστικού Δέκτη (Receiver Operating Characteristic)
- S1: Πρωτοταγής Σωματοαισθητικός Φλοιός
- SHAP: SHapley Additive exPlanations (Προσθετικές Επεξηγήσεις Shapley)
- SL: Source Localization (Εντοπισμός Πηγής)
- sLORETA: Τυποποιημένη LORETA
- SNR: Λόγος Σήματος προς Θόρυβο
- SS-FIS: Σταθερού Σταθερού Διαστήματος Εξομάλυνσης (Φίλτρο Kalman)
- SS-KF: Φίλτρο Kalman Σταθερής Κατάστασης
- SVM: Υποστηρικτικές Μηχανές Διανυσμάτων
- TFR: Time-Frequency Representation (Αναπαράσταση Χρόνου-Συχνότητας) (Χρησιμοποιείται ως TFR στο κείμενο, η πλήρης ελληνική φράση είναι Ανάλυση Χρόνου-Συχνότητας)
- ΤΝ: Τεχνητή Νοημοσύνη
- **TPR:** Ρυθμός Αληθώς Θετικών (True Positive Rate)
- ΧΑΙ: Εξηγήσιμη Τεχνητή Νοημοσύνη
- **Fp**, **F**, **T**, **C**, **P**, **O**, **Z**: (Συμβολισμοί θέσεων ηλεκτροδίων ΗΕΓ)
- AF, FC, FT, CP, TP, PO: (Συμβολισμοί ενδιάμεσων θέσεων ηλεκτροδίων ΗΕΓ)

## **Glossary of Abbreviations (English)**

- **AI:** Artificial Intelligence
- **Ag/AgCl:** Silver/Silver-Chloride
- AUC / AUROC: Area Under the ROC (Receiver Operating Characteristic) Curve
- **BAE:** Bayesian Approximation Error
- BCI(s) ( $\Delta \Delta EY$ ): Brain-Computer Interface(s)
- **BEM:** Boundary Element Method
- **CBAM:** Convolutional Block Attention Module
- CG-FEM: Continuous Galerkin-FEM (Finite Element Method)
- **CNN(s):** Convolutional Neural Network(s)
- **ConvDip:** (Name of a specific Convolutional Neural Network)
- CSF: Cerebrospinal Fluid
- **DeepLIFT:** Deep Learning Important FeaTures
- **DL (BM):** Deep Learning
- **DST-DAE:** (Name of a specific Denoising Autoencoder)
- **dw-MNE:** depth-weighted Minimum Norm Estimate
- ECDs: Equivalent Current Dipoles
- ECG/EKG: Electrocardiography
- **EEG:** Electroencephalography
- eLORETA: exact Low-Resolution Electromagnetic Tomography
- EMG: Electromyography
- **EOG:** Electrooculogram
- **ERPs:** Event-Related Potentials
- **ESL:** EEG Source Localization
- FC: Fully Connected Layer
- **FEM:** Finite Element Method
- **FIR:** Finite Impulse Response [filter]
- fMRI: functional Magnetic Resonance Imaging
- Grad-CAM: Gradient-weighted Class Activation Mapping
- HD-ESI: High-Density EEG Source Imaging
- ICA: Independent Component Analysis
- IG: Integrated Gradients
- **IIR:** Infinite Impulse Response [filter]
- **Iz:** (Inion electrode position)
- LCMV: Linearly Constrained Minimum Variance
- LD-ESI: Low-Density EEG Source Imaging
- **LFM:** Leadfield Matrix
- LIME: Local Interpretable Model-Agnostic Explanations
- LORETA: Low-Resolution Electromagnetic Tomography [page 35, page 62]
- LSTM: Long Short-Term Memory [units]
- M1 (Motor Cortex): Primary Motor Cortex
- M1, M2 (Electrodes): Mastoid Electrodes
- MEG (MEG): Magnetoencephalography
- MI (KA): Motor Imagery
- **MNE:** Minimum Norm Estimate
- MNI: Montreal Neurological Institute
- MRI (MT): Magnetic Resonance Imaging
- PCANet: Principal Component Analysis Network
- **PM:** Projective Method
- **PPV:** Positive Predictive Value
- **PSD:** Power Spectral Density
- **ROIs:** Regions of Interest
- ROC: Receiver Operating Characteristic [curve]
- S1: Primary Somatosensory Cortex
- SHAP: SHapley Additive exPlanations
- SL: Source Localization
- **sLORETA:** standardized LORETA
- SMA: Supplementary Motor Area
- **SNR:** Signal-to-Noise Ratio
- SS-FIS: Steady-State Fixed-Interval Smoother [Kalman filter]
- SS-KF: Steady-State Kalman Filter
- SVM: Support Vector Machine
- TFR: Time-Frequency Representation
- **TPR:** True Positive Rate
- XAI: Explainable Artificial Intelligence
- **Fp**, **F**, **T**, **C**, **P**, **O**, **Z**: (EEG electrode position labels)
- AF, FC, FT, CP, TP, PO: (Intermediate EEG electrode position labels)

## Βιβλιογραφία

- [1] C. T. Tan, 'Electroencephalography in Clinical Practice', *Ann Indian Acad Neurol*, vol. 21, no. 3, p. 236, 2018, doi: 10.4103/aian.AIAN\_342\_18.
- [2] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, 'Brain–computer interfaces for communication and control', *Clinical Neurophysiology*, vol. 113, no. 6, pp. 767–791, Jun. 2002, doi: 10.1016/S1388-2457(02)00057-3.
- [3] H. Cecotti and A. Graser, 'Convolutional Neural Networks for P300 Detection with Application to Brain-Computer Interfaces', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 433–445, Mar. 2011, doi: 10.1109/TPAMI.2010.125.
- [4] Z. C. Lipton, 'The Mythos of Model Interpretability', Mar. 06, 2017, *arXiv*: arXiv:1606.03490. doi: 10.48550/arXiv.1606.03490.
- [5] B. Goodman and S. Flaxman, 'European Union Regulations on Algorithmic Decision Making and a "Right to Explanation", *AI Magazine*, vol. 38, no. 3, pp. 50–57, 2017, doi: 10.1609/aimag.v38i3.2741.
- [6] Y. Yang, Q. M. J. Wu, W.-L. Zheng, and B.-L. Lu, 'EEG-Based Emotion Recognition Using Hierarchical Network With Subnetwork Nodes', *IEEE Transactions on Cognitive* and Developmental Systems, vol. 10, no. 2, pp. 408–419, Jun. 2018, doi: 10.1109/TCDS.2017.2685338.
- [7] G. Maral, VSAT Networks, 2nd edition. Chichester: Wiley, 2004.
- [8] C. Tallon-Baudry, O. Bertrand, C. Delpuech, and J. Pernier, 'Oscillatory γ-Band (30–70 Hz) Activity Induced by a Visual Search Task in Humans', J. Neurosci., vol. 17, no. 2, pp. 722–734, Jan. 1997, doi: 10.1523/JNEUROSCI.17-02-00722.1997.
- [9] I. Zorzos, I. Kakkos, E. M. Ventouras, and G. K. Matsopoulos, 'Advances in Electrical Source Imaging: A Review of the Current Approaches, Applications and Challenges', *Signals*, vol. 2, no. 3, 2021, doi: 10.3390/signals2030024.
- [10] C. M. Michel, M. M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R. Grave de Peralta, 'EEG source imaging', *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2195– 2222, Oct. 2004, doi: 10.1016/j.clinph.2004.06.001.
- [11] P. Berg and M. Scherg, 'A fast method for forward computation of multiple-shell spherical head models', *Electroencephalography and Clinical Neurophysiology*, vol. 90, no. 1, pp. 58–64, Jan. 1994, doi: 10.1016/0013-4694(94)90113-9.
- [12] M. S. Hamalainen and J. Sarvas, 'Realistic conductivity geometry model of the human head for interpretation of neuromagnetic data', *IEEE Transactions on Biomedical Engineering*, vol. 36, no. 2, pp. 165–171, Feb. 1989, doi: 10.1109/10.16463.
- [13] C. H. Wolters, A. Anwander, X. Tricoche, D. Weinstein, M. A. Koch, and R. S. MacLeod, 'Influence of tissue conductivity anisotropy on EEG/MEG field and return current computation in a realistic head model: a simulation and visualization study using high-resolution finite element modeling', *Neuroimage*, vol. 30, no. 3, pp. 813–826, Apr. 2006, doi: 10.1016/j.neuroimage.2005.10.014.
- [14] R. Pohlmeier, H. Buchner, G. Knoll, A. RienÄcker, R. Beckmann, and J. Pesch, 'The influence of skull-conductivity misspecification on inverse source localization in realistically shaped finite element head models', *Brain Topogr*, vol. 9, no. 3, pp. 157– 162, Mar. 1997, doi: 10.1007/BF01190384.
- [15] D. Weinstein, L. Zhukov, and C. Johnson, 'Lead-field Bases for Electroencephalography Source Imaging', *Annals of Biomedical Engineering*, vol. 28, no. 9, pp. 1059–1065, Sep. 2000, doi: 10.1114/1.1310220.

- [16] F. Grandori and M. Hoke, *Auditory Evoked Magnetic Fields and Electric Potentials*, 1st edition. Basel: S Karger Ag, 1990.
- [17] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa, 'Magnetoencephalography---theory, instrumentation, and applications to noninvasive studies of the working human brain', *Rev. Mod. Phys.*, vol. 65, no. 2, pp. 413–497, Apr. 1993, doi: 10.1103/RevModPhys.65.413.
- [18] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann, 'Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain', *International Journal of Psychophysiology*, vol. 18, no. 1, pp. 49–65, Oct. 1994, doi: 10.1016/0167-8760(84)90014-X.
- [19] R. D. Pascual-Marqui, 'Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details', *Methods Find Exp Clin Pharmacol*, vol. 24 Suppl D, pp. 5–12, 2002.
- [20] R. D. Pascual-Marqui, 'Discrete, 3D distributed, linear imaging methods of electric neuronal activity. Part 1: exact, zero error localization', Oct. 18, 2007, arXiv: arXiv:0710.3341. doi: 10.48550/arXiv.0710.3341.
- [21] B. D. Van Veen, W. Van Drongelen, M. Yuchtman, and A. Suzuki, 'Localization of brain electrical activity via linearly constrained minimum variance spatial filtering', *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 867–880, Sep. 1997, doi: 10.1109/10.623056.
- [22] C. M. Michel and B. He, 'Chapter 6 EEG source localization', in *Handbook of Clinical Neurology*, vol. 160, K. H. Levin and P. Chauvel, Eds., Elsevier, 2019, pp. 85–101. doi: 10.1016/B978-0-444-64032-1.00006-0.
- [23] J. Rahul, D. Sharma, L. D. Sharma, U. Nanda, and A. K. Sarkar, 'A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning', *Front. Hum. Neurosci.*, vol. 18, Feb. 2024, doi: 10.3389/fnhum.2024.1347082.
- [24] Y. Pourasad, E. Zarouri, M. Salemizadeh Parizi, and A. Salih Mohammed, 'Presentation of Novel Architecture for Diagnosis and Identifying Breast Cancer Location Based on Ultrasound Images Using Machine Learning', *Diagnostics (Basel)*, vol. 11, no. 10, p. 1870, Oct. 2021, doi: 10.3390/diagnostics11101870.
- [25] Y. LeCun, Y. Bengio, and G. Hinton, 'Deep learning', *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [26] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, 'Dropout: a simple way to prevent neural networks from overfitting', *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [27] S. Ioffe and C. Szegedy, 'Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift', in *Proceedings of the 32nd International Conference* on Machine Learning, F. Bach and D. Blei, Eds., in Proceedings of Machine Learning Research, vol. 37. Lille, France: PMLR, Jul. 2015, pp. 448–456. [Online]. Available: https://proceedings.mlr.press/v37/ioffe15.html
- [28] S. A. Hicks *et al.*, 'On evaluation metrics for medical applications of artificial intelligence', *Sci Rep*, vol. 12, p. 5979, Apr. 2022, doi: 10.1038/s41598-022-09954-8.
- [29] A. P. Bradley, 'The use of the area under the ROC curve in the evaluation of machine learning algorithms', *Pattern Recognition*, vol. 30, no. 7, pp. 1145–1159, Jul. 1997, doi: 10.1016/S0031-3203(96)00142-2.
- [30] S. M. Lundberg and S.-I. Lee, 'A Unified Approach to Interpreting Model Predictions', in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S.

Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

- [31] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?": Explaining the Predictions of Any Classifier', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, Aug. 2016, pp. 1135–1144. doi: 10.1145/2939672.2939778.
- [32] M. Sundararajan, A. Taly, and Q. Yan, 'Axiomatic attribution for deep networks', in Proceedings of the 34th International Conference on Machine Learning - Volume 70, in ICML'17. Sydney, NSW, Australia: JMLR.org, Aug. 2017, pp. 3319–3328.
- [33] L. S. Shapley and others, 'A value for n-person games', 1953.
- [34] H. Alsuradi, W. Park, and M. Eid, 'Explainable Classification of EEG Data for an Active Touch Task Using Shapley Values', in *HCI International 2020 - Late Breaking Papers: Multimodality and Intelligence*, C. Stephanidis, M. Kurosu, H. Degen, and L. Reinerman-Jones, Eds., Cham: Springer International Publishing, 2020, pp. 406–416. doi: 10.1007/978-3-030-60117-1\_30.
- [35] M. Slimani, H. Znazen, N. L. Bragazzi, M. S. Zguira, and D. Tod, 'The Effect of Mental Fatigue on Cognitive and Aerobic Performance in Adolescent Active Endurance Athletes: Insights from a Randomized Counterbalanced, Cross-Over Trial', *J Clin Med*, vol. 7, no. 12, p. 510, Dec. 2018, doi: 10.3390/jcm7120510.
- [36] V. Pergher, N. Vanbilsen, and M. Van Hulle, 'The Effect of Mental Fatigue and Gender on Working Memory Performance during Repeated Practice by Young and Older Adults', *Neural Plasticity*, vol. 2021, p. e6612805, Oct. 2021, doi: 10.1155/2021/6612805.
- [37] A. Welp, L. L. Meier, and T. Manser, 'Emotional exhaustion and workload predict clinician-rated and objective patient safety', *Frontiers in Psychology*, vol. 5, 2015, Accessed: Jun. 15, 2022. [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyg.2014.01573
- [38] M. A. S. Boksem, T. F. Meijman, and M. M. Lorist, 'Effects of mental fatigue on attention: An ERP study', *Cognitive Brain Research*, vol. 25, no. 1, pp. 107–116, Sep. 2005, doi: 10.1016/j.cogbrainres.2005.04.011.
- [39] J. DeLuca, H. M. Genova, F. G. Hillary, and G. Wylie, 'Neural correlates of cognitive fatigue in multiple sclerosis using functional MRI', *Journal of the Neurological Sciences*, vol. 270, no. 1, pp. 28–39, Jul. 2008, doi: 10.1016/j.jns.2008.01.018.
- [40] G. Li *et al.*, 'The impact of mental fatigue on brain activity: a comparative study both in resting state and task state using EEG', *BMC Neuroscience*, vol. 21, no. 1, p. 20, May 2020, doi: 10.1186/s12868-020-00569-1.
- [41] G. Shou, L. Ding, and D. Dasari, 'Probing neural activations from continuous EEG in a real-world task: Time-frequency independent component analysis', *Journal of Neuroscience Methods*, vol. 209, no. 1, pp. 22–34, Jul. 2012, doi: 10.1016/j.jneumeth.2012.05.022.
- [42] D. Dasari, G. Shou, and L. Ding, 'ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task', *Frontiers in Neuroscience*, vol. 11, 2017, Accessed: Dec. 30, 2022. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2017.00297

- [43] Y. Liu, Z. Lan, H. H. G. Khoo, K. H. H. Li, O. Sourina, and W. Mueller-Wittig, 'EEGbased Evaluation of Mental Fatigue Using Machine Learning Algorithms', in 2018 International Conference on Cyberworlds (CW), Oct. 2018, pp. 276–279. doi: 10.1109/CW.2018.00056.
- [44] Z. Gao *et al.*, 'EEG-Based Spatio–Temporal Convolutional Neural Network for Driver Fatigue Evaluation', *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2755–2763, Sep. 2019, doi: 10.1109/TNNLS.2018.2886414.
- [45] V. P. Balam, V. U. Sameer, and S. Chinara, 'Automated classification system for drowsiness detection using convolutional neural network and electroencephalogram', *IET Intelligent Transport Systems*, vol. 15, no. 4, pp. 514–524, 2021, doi: 10.1049/itr2.12041.
- [46] S. Khessiba, A. G. Blaiech, K. Ben Khalifa, A. Ben Abdallah, and M. H. Bedoui, 'Innovative deep learning models for EEG-based vigilance detection', *Neural Computing and Applications*, vol. 33, no. 12, pp. 6921–6937, Jun. 2021, doi: 10.1007/s00521-020-05467-5.
- [47] Q. Zhang and S. Zhu, 'Visual interpretability for deep learning: a survey', Frontiers Inf Technol Electronic Eng, vol. 19, no. 1, pp. 27–39, Jan. 2018, doi: 10.1631/FITEE.1700808.
- [48] I. Kakkos *et al.*, 'EEG fingerprints of task-independent mental workload discrimination', *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2021, doi: 10.1109/JBHI.2021.3085131.
- [49] G. N. Dimitrakopoulos, I. Kakkos, N. V. Thakor, A. Bezerianos, and null Yu Sun, 'A mental fatigue index based on regression using mulitband EEG features with application in simulated driving', *Conf Proc IEEE Eng Med Biol Soc*, vol. 2017, pp. 3220–3223, 2017, doi: 10.1109/EMBC.2017.8037542.
- [50] S. T. Miloulis *et al.*, 'Stimulus Effects on Subject-Specific BCI Classification Training using Motor Imagery', in 2021 International Conference on e-Health and Bioengineering (EHB), Nov. 2021, pp. 1–4. doi: 10.1109/EHB52898.2021.9657538.
- [51] A. Delorme and S. Makeig, 'EEGLAB: an open source toolbox for analysis of singletrial EEG dynamics including independent component analysis', *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004, doi: 10.1016/j.jneumeth.2003.10.009.
- [52] S. Morales and M. E. Bowers, 'Time-frequency analysis methods and their application in developmental EEG data', *Developmental Cognitive Neuroscience*, vol. 54, p. 101067, Apr. 2022, doi: 10.1016/j.dcn.2022.101067.
- [53] S. Kiebel, J. Kilner, and K. Friston, 'CHAPTER 16 Hierarchical models for EEG and MEG', in *Statistical Parametric Mapping*, K. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny, Eds., London: Academic Press, 2007, pp. 211–220. doi: 10.1016/B978-012372560-8/50016-4.
- [54] C. D'Avanzo, V. Tarantino, P. Bisiacchi, and G. Sparacino, 'A wavelet Methodology for EEG Time-frequency Analysis in a Time Discrimination Task', *Int. J. Bioelectromagn.*, vol. 11, no. 4, pp. 185–188, 2009.
- [55] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, 'FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data', *Computational Intelligence and Neuroscience*, vol. 2011, p. e156869, Dec. 2010, doi: 10.1155/2011/156869.
- [56] T. Tsoneva, D. Baldo, V. Lema, and G. Garcia-Molina, 'EEG-rhythm dynamics during a 2-back working memory task and performance', in 2011 Annual International

*Conference of the IEEE Engineering in Medicine and Biology Society*, Aug. 2011, pp. 3828–3831. doi: 10.1109/IEMBS.2011.6090952.

- [57] Q. Sellat, S. K. Bisoy, and R. Priyadarshini, 'Chapter 10 Semantic segmentation for self-driving cars using deep learning: a survey', in *Cognitive Big Data Intelligence with a Metaheuristic Approach*, S. Mishra, H. K. Tripathy, P. K. Mallick, A. K. Sangaiah, and G.-S. Chae, Eds., in Cognitive Data Science in Sustainable Computing., Academic Press, 2022, pp. 211–238. doi: 10.1016/B978-0-323-85117-6.00002-9.
- [58] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization', Jan. 29, 2017, arXiv: arXiv:1412.6980. doi: 10.48550/arXiv.1412.6980.
- [59] A. Paszke et al., 'PyTorch: An Imperative Style, High-Performance Deep Learning Library', in Advances in Neural Information Processing Systems 32, Curran Associates, Inc., 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
- [60] Y. Ma et al., 'Driving Fatigue Detection from EEG Using a Modified PCANet Method', Computational Intelligence and Neuroscience, vol. 2019, p. e4721863, Jul. 2019, doi: 10.1155/2019/4721863.
- [61] A. Abidi, K. Ben Khalifa, R. Ben Cheikh, C. A. Valderrama Sakuyama, and M. H. Bedoui, 'Automatic Detection of Drowsiness in EEG Records Based on Machine Learning Approaches', *Neural Process Lett*, vol. 54, no. 6, pp. 5225–5249, Dec. 2022, doi: 10.1007/s11063-022-10858-x.
- [62] A. Shrikumar, P. Greenside, and A. Kundaje, 'Learning Important Features Through Propagating Activation Differences', in *Proceedings of the 34th International Conference on Machine Learning*, PMLR, Jul. 2017, pp. 3145–3153. Accessed: Dec. 30, 2022. [Online]. Available: https://proceedings.mlr.press/v70/shrikumar17a.html
- [63] B. J. Roach and D. H. Mathalon, 'Event-Related EEG Time-Frequency Analysis: An Overview of Measures and An Analysis of Early Gamma Band Phase Locking in Schizophrenia', *Schizophrenia Bulletin*, vol. 34, no. 5, pp. 907–926, Sep. 2008, doi: 10.1093/schbul/sbn093.
- [64] B. Venkata Phanikrishna, A. Jaya Prakash, and C. Suchismitha, 'Deep Review of Machine Learning Techniques on Detection of Drowsiness Using EEG Signal', *null*, pp. 1–16, May 2021, doi: 10.1080/03772063.2021.1913070.
- [65] S. Scheurer, S. Tedesco, B. O'Flynn, and K. N. Brown, 'Comparing Person-Specific and Independent Models on Subject-Dependent and Independent Human Activity Recognition Performance', *Sensors (Basel)*, vol. 20, no. 13, p. 3647, Jun. 2020, doi: 10.3390/s20133647.
- [66] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, 'Edge Computing: Vision and Challenges', *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.
- [67] G. N. Dimitrakopoulos *et al.*, 'Functional Connectivity Analysis of Mental Fatigue Reveals Different Network Topological Alterations Between Driving and Vigilance Tasks', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 4, pp. 740–749, Apr. 2018, doi: 10.1109/TNSRE.2018.2791936.
- [68] A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, 'Estimating workload using EEG spectral power and ERPs in the nback task', *J Neural Eng*, vol. 9, no. 4, p. 045008, Aug. 2012, doi: 10.1088/1741-2560/9/4/045008.
- [69] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, 'Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental

workload, fatigue and drowsiness', *Neuroscience & Biobehavioral Reviews*, vol. 44, pp. 58–75, Jul. 2014, doi: 10.1016/j.neubiorev.2012.10.003.

- [70] M. Crespo-Garcia, D. Pinal, J. L. Cantero, F. Díaz, M. Zurrón, and M. Atienza, 'Working memory processes are mediated by local and long-range synchronization of alpha oscillations', *J Cogn Neurosci*, vol. 25, no. 8, pp. 1343–1357, Aug. 2013, doi: 10.1162/jocn\_a\_00379.
- [71] K. Fukuda, I. Mance, and E. K. Vogel, 'α Power Modulation and Event-Related Slow Wave Provide Dissociable Correlates of Visual Working Memory', J. Neurosci., vol. 35, no. 41, pp. 14009–14016, Oct. 2015, doi: 10.1523/JNEUROSCI.5003-14.2015.
- [72] C. M. Krause, M. Pesonen, and H. Hämäläinen, 'Brain oscillatory 4-30 Hz electroencephalogram responses in adolescents during a visual memory task', *Neuroreport*, vol. 21, no. 11, pp. 767–771, Aug. 2010, doi: 10.1097/WNR.0b013e32833bfcbb.
- [73] G. N. Dimitrakopoulos *et al.*, 'Task-Independent Mental Workload Classification Based Upon Common Multiband EEG Cortical Connectivity', *IEEE Trans Neural Syst Rehabil Eng*, vol. 25, no. 11, pp. 1940–1949, 2017, doi: 10.1109/TNSRE.2017.2701002.
- [74] K. S. Chiew and T. S. Braver, 'Context processing and cognitive control: From gating models to dual mechanisms', in *The Wiley handbook of cognitive control*, Hoboken, NJ, US: Wiley Blackwell, 2017, pp. 143–166. doi: 10.1002/9781118920497.ch9.
- [75] A. Ardestani, W. Shen, F. Darvas, A. W. Toga, and J. M. Fuster, 'Modulation of Frontoparietal Neurovascular Dynamics in Working Memory', *Journal of Cognitive Neuroscience*, vol. 28, no. 3, pp. 379–401, Dec. 2015, doi: 10.1162/jocn\_a\_00903.
- [76] I. Stancin, M. Cifrek, and A. Jovic, 'A Review of EEG Signal Features and Their Application in Driver Drowsiness Detection Systems', *Sensors*, vol. 21, no. 11, Art. no. 11, Jan. 2021, doi: 10.3390/s21113786.
- [77] U. Mitzdorf, 'Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and EEG phenomena', *Physiological Reviews*, vol. 65, no. 1, pp. 37–100, Jan. 1985, doi: 10.1152/physrev.1985.65.1.37.
- [78] H. Helmholtz, 'Ueber einige Gesetze der Vertheilung elektrischer Ströme in körperlichen Leitern mit Anwendung auf die thierisch-elektrischen Versuche', Annalen der Physik, vol. 165, no. 6, pp. 211–233, Jan. 1853, doi: 10.1002/andp.18531650603.
- [79] S. T. Hansen *et al.*, 'Unmixing Oscillatory Brain Activity by EEG Source Localization and Empirical Mode Decomposition', *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–15, Mar. 2019, doi: 10.1155/2019/5618303.
- [80] K. Mahjoory, V. V. Nikulin, L. Botrel, K. Linkenkaer-Hansen, M. M. Fato, and S. Haufe, 'Consistency of EEG source localization and connectivity estimates', *NeuroImage*, vol. 152, pp. 590–601, May 2017, doi: 10.1016/j.neuroimage.2017.02.076.
- [81] M. Stropahl, A.-K. R. Bauer, S. Debener, and M. G. Bleichner, 'Source-Modeling Auditory Processes of EEG Data Using EEGLAB and Brainstorm', *Front. Neurosci.*, vol. 12, 2018, doi: 10.3389/fnins.2018.00309.
- [82] Donald L. Schomer, Fernando H. Lopes da Silva, Christoph M. Michel, and Bin He, 'EEG Mapping and Source Imaging', Oxford, UK: Oxford University Press, 2017. doi: 10.1093/med/9780190228484.003.0045.
- [83] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The neurophysics of EEG*, 2nd ed. New York: Oxford University Press, 2006. doi: 10.1093/acprof:oso/9780195050387.001.0001.
- [84] C. Phillips, M. D. Rugg, and K. J. Friston, 'Anatomically Informed Basis Functions for EEG Source Localization: Combining Functional and Anatomical Constraints',

*NeuroImage*, vol. 16, no. 3, Part A, pp. 678–695, Jul. 2002, doi: 10.1006/nimg.2002.1143.

- [85] 'Frontiers | EEG Source Imaging: A Practical Review of the Analysis Steps | Neurology'. Accessed: Feb. 25, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2019.00325/full
- [86] S. Cui et al., 'EEG source localization using spatio-temporal neural network', China Communications, vol. 16, no. 7, pp. 131–143, Jul. 2019, doi: 10.23919/JCC.2019.07.011.
- [87] L. Hecker, R. Rupprecht, L. Tebartz Van Elst, and J. Kornmeier, 'ConvDip: A Convolutional Neural Network for Better EEG Source Imaging', *Front. Neurosci.*, vol. 15, Jun. 2021, doi: 10.3389/fnins.2021.569918.
- [88] F. Costa, H. Batatia, T. Oberlin, C. D'Giano, and J.-Y. Tourneret, 'Bayesian EEG source localization using a structured sparsity prior', *NeuroImage*, vol. 144, pp. 142–152, Jan. 2017, doi: 10.1016/j.neuroimage.2016.08.064.
- [89] R. Kar, A. Konar, A. Chakraborty, B. S. Bhattacharya, and A. K. Nagar, 'EEG source localization by memory network analysis of subjects engaged in perceiving emotions from facial expressions', in 2015 International Joint Conference on Neural Networks (IJCNN), Jul. 2015, pp. 1–8. doi: 10.1109/IJCNN.2015.7280705.
- [90] J. P. Ary, S. A. Klein, and D. H. Fender, 'Location of Sources of Evoked Scalp Potentials: Corrections for Skull and Scalp Thicknesses', *IEEE Transactions on Biomedical Engineering*, vol. BME-28, no. 6, pp. 447–452, Jun. 1981, doi: 10.1109/TBME.1981.324817.
- [91] J.-H. Cho, J. Vorwerk, C. H. Wolters, and T. R. Knösche, 'Influence of the head model on EEG and MEG source connectivity analyses', *NeuroImage*, vol. 110, pp. 60–77, Apr. 2015, doi: 10.1016/j.neuroimage.2015.01.043.
- [92] J. Vorwerk, J.-H. Cho, S. Rampp, H. Hamer, T. R. Knösche, and C. H. Wolters, 'A guideline for head volume conductor modeling in EEG and MEG', *NeuroImage*, vol. 100, pp. 590–607, Oct. 2014, doi: 10.1016/j.neuroimage.2014.06.040.
- [93] H. I. Saleheen and K. T. Ng, 'New finite difference formulations for general inhomogeneous anisotropic bioelectric problems', *IEEE Transactions on Biomedical Engineering*, vol. 44, no. 9, pp. 800–809, Sep. 1997, doi: 10.1109/10.623049.
- [94] M. Fuchs, M. Wagner, and J. Kastner, 'Boundary element method volume conductor models for EEG source reconstruction', *Clinical Neurophysiology*, vol. 112, no. 8, pp. 1400–1407, Aug. 2001, doi: 10.1016/S1388-2457(01)00589-2.
- [95] M. Rullmann, A. Anwander, M. Dannhauer, S. K. Warfield, F. H. Duffy, and C. H. Wolters, 'EEG source analysis of epileptiform activity using a 1 mm anisotropic hexahedra finite element head model', *NeuroImage*, vol. 44, no. 2, pp. 399–410, Jan. 2009, doi: 10.1016/j.neuroimage.2008.09.009.
- [96] H.-S. Wong, B. Ma, Z. Yu, P. F. Yeung, and H. H. S. Ip, '3-D Head Model Retrieval Using a Single Face View Query', *IEEE Transactions on Multimedia*, vol. 9, no. 5, pp. 1026–1036, Aug. 2007, doi: 10.1109/TMM.2007.898915.
- [97] T. Nara and S. Ando, 'A projective method for an inverse source problem of the Poisson equation', *Inverse Problems*, vol. 19, no. 2, pp. 355–369, Feb. 2003, doi: 10.1088/0266-5611/19/2/307.
- [98] V. Saase, H. Wenz, T. Ganslandt, C. Groden, and M. E. Maros, 'Simple statistical methods for unsupervised brain anomaly detection on MRI are competitive to deep learning methods', *arXiv:2011.12735 [cs, eess]*, Nov. 2020, Accessed: Apr. 25, 2021. [Online]. Available: http://arxiv.org/abs/2011.12735

- [99] J. Vorwerk, R. Oostenveld, M. C. Piastra, L. Magyari, and C. H. Wolters, 'The FieldTrip-SimBio pipeline for EEG forward solutions', *BioMedical Engineering OnLine*, vol. 17, no. 1, p. 37, Mar. 2018, doi: 10.1186/s12938-018-0463-y.
- [100] Y. Huang, L. C. Parra, and S. Haufe, 'The New York Head—A precise standardized volume conductor model for EEG source localization and tES targeting', *NeuroImage*, vol. 140, pp. 150–162, Oct. 2016, doi: 10.1016/j.neuroimage.2015.12.019.
- [101] L. Beltrachini, 'A Finite Element Solution of the Forward Problem in EEG for Multipolar Sources', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 3, pp. 368–377, Mar. 2019, doi: 10.1109/TNSRE.2018.2886638.
- [102] G. Birot *et al.*, 'Head model and electrical source imaging: A study of 38 epileptic patients', *NeuroImage: Clinical*, vol. 5, pp. 77–83, Jan. 2014, doi: 10.1016/j.nicl.2014.06.005.
- [103] C. Engwer, J. Vorwerk, J. Ludewig, and C. H. Wolters, 'A Discontinuous Galerkin Method to Solve the EEG Forward Problem Using the Subtraction Approach', *SIAM J. Sci. Comput.*, vol. 39, no. 1, pp. B138–B164, Jan. 2017, doi: 10.1137/15M1048392.
- [104] J. Vorwerk, C. Engwer, S. Pursiainen, and C. H. Wolters, 'A Mixed Finite Element Method to Solve the EEG Forward Problem', *IEEE Transactions on Medical Imaging*, vol. 36, no. 4, pp. 930–941, Apr. 2017, doi: 10.1109/TMI.2016.2624634.
- [105] L. Beltrachini, 'The analytical subtraction approach for solving the forward problem in EEG', *Journal of Neural Engineering*, vol. 16, no. 5, p. 056029, Sep. 2019, doi: 10.1088/1741-2552/ab2694.
- [106] T. Miinalainen *et al.*, 'A realistic, accurate and fast source modeling approach for the EEG forward problem', *NeuroImage*, vol. 184, pp. 56–67, Jan. 2019, doi: 10.1016/j.neuroimage.2018.08.054.
- [107] B. Vanrumste, G. Van Hoey, R. Van de Walle, M. R. P. D'Havè, I. A. Lemahieu, and P. A. J. M. Boon, 'The Validation of the Finite Difference Method and Reciprocity for Solving the Inverse Problem in EEG Dipole Source Analysis', *Brain Topography*, vol. 14, no. 2, pp. 83–92, Dec. 2001, doi: 10.1023/A:1012909511833.
- [108] E. Cuartas Morales, C. D. Acosta-Medina, G. Castellanos-Dominguez, and D. Mantini, 'A Finite-Difference Solution for the EEG Forward Problem in Inhomogeneous Anisotropic Media', *Brain Topography*, vol. 32, no. 2, pp. 229–239, Mar. 2019, doi: 10.1007/s10548-018-0683-2.
- [109] A. Pillain, L. Rahmouni, and F. Andriulli, 'Handling anisotropic conductivities in the EEG forward problem with a symmetric formulation', *Physics in Medicine & Biology*, vol. 64, no. 3, p. 035022, Feb. 2019, doi: 10.1088/1361-6560/aafaaf.
- [110] V. Montes-Restrepo, P. van Mierlo, G. Strobbe, S. Staelens, S. Vandenberghe, and H. Hallez, 'Influence of Skull Modeling Approaches on EEG Source Localization', *Brain Topography*, vol. 27, no. 1, pp. 95–111, Jan. 2014, doi: 10.1007/s10548-013-0313-y.
- [111] V. Rimpiläinen, A. Koulouri, F. Lucka, J. P. Kaipio, and C. H. Wolters, 'Improved EEG source localization with Bayesian uncertainty modelling of unknown skull conductivity', *NeuroImage*, vol. 188, pp. 252–260, Mar. 2019, doi: 10.1016/j.neuroimage.2018.11.058.
- [112] DH. FENDER, 'Source localization of brain electrical activity', Handbook of electroencephalography and clinical neurophysiology, pp. 355–403, 1987.
- [113] R. Grech *et al.*, 'Review on solving the inverse problem in EEG source analysis', *Journal of NeuroEngineering and Rehabilitation*, vol. 5, no. 1, p. 25, Nov. 2008, doi: 10.1186/1743-0003-5-25.

- [114] M. S. Hämäläinen and R. J. Ilmoniemi, 'Interpreting magnetic fields of the brain: minimum norm estimates', *Med. Biol. Eng. Comput.*, vol. 32, no. 1, pp. 35–42, Jan. 1994, doi: 10.1007/BF02512476.
- [115] M. Fuchs, M. Wagner, T. Köhler, and H.-A. Wischmann, 'Linear and Nonlinear Current Density Reconstructions', *Journal of Clinical Neurophysiology*, vol. 16, no. 3, pp. 267–295, May 1999.
- [116] F.-H. Lin, T. Witzel, S. P. Ahlfors, S. M. Stufflebeam, J. W. Belliveau, and M. S. Hämäläinen, 'Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates', *NeuroImage*, vol. 31, no. 1, pp. 160–171, May 2006, doi: 10.1016/j.neuroimage.2005.11.054.
- [117] R. D. Pascual-Marqui et al., 'Assessing interactions in the brain with exact lowresolution electromagnetic tomography', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 369, no. 1952, pp. 3768–3784, Oct. 2011, doi: 10.1098/rsta.2011.0081.
- [118] J. Vrba and S. E. Robinson, 'Signal Processing in Magnetoencephalography', *Methods*, vol. 25, no. 2, pp. 249–271, Oct. 2001, doi: 10.1006/meth.2001.1238.
- [119] L. Hecker, R. Rupprecht, L. Tebartz van Elst, and J. Kornmeier, 'ConvDip: A convolutional neural network for better M/EEG Source Imaging', *bioRxiv*, p. 2020.04.09.033506, Jan. 2020, doi: 10.1101/2020.04.09.033506.
- [120] G. Huang et al., 'Electromagnetic Source Imaging via a Data-Synthesis-Based Denoising Autoencoder', arXiv:2010.12876 [cs, eess], Jan. 2021, Accessed: Feb. 09, 2021. [Online]. Available: http://arxiv.org/abs/2010.12876
- [121] D. P. Wipf, J. P. Owen, H. T. Attias, K. Sekihara, and S. S. Nagarajan, 'Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG', *NeuroImage*, vol. 49, no. 1, pp. 641–655, Jan. 2010, doi: 10.1016/j.neuroimage.2009.06.083.
- [122] E. Pirondini *et al.*, 'Computationally Efficient Algorithms for Sparse, Dynamic Solutions to the EEG Source Localization Problem', *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 6, pp. 1359–1372, Jun. 2018, doi: 10.1109/TBME.2017.2739824.
- [123] E. Ebrahimzadeh *et al.*, 'Epilepsy Presurgical Evaluation of Patients with Complex Source Localization by a Novel Component-Based EEG-fMRI Approach', *Iran J Radiol*, vol. 16, no. Special Issue, Art. no. Special Issue, Nov. 2019, doi: 10.5812/iranjradiol.99134.
- [124] C. Plummer, S. J. Vogrin, W. P. Woods, M. A. Murphy, M. J. Cook, and D. T. J. Liley, 'Interictal and ictal source localization for epilepsy surgery using high-density EEG with MEG: a prospective long-term study', *Brain*, vol. 142, no. 4, pp. 932–951, Apr. 2019, doi: 10.1093/brain/awz015.
- [125] P. Nemtsas *et al.*, 'Source localization of ictal epileptic activity based on high-density scalp EEG data', *Epilepsia*, vol. 58, no. 6, pp. 1027–1036, 2017, doi: https://doi.org/10.1111/epi.13749.
- [126] M. Centeno *et al.*, 'Combined electroencephalography–functional magnetic resonance imaging and electrical source imaging improves localization of pediatric focal epilepsy', *Annals of Neurology*, vol. 82, no. 2, pp. 278–287, 2017, doi: https://doi.org/10.1002/ana.25003.
- [127] C. M. Michel and B. He, *EEG Mapping and Source Imaging*. Oxford University Press.Accessed:Feb.13,2021.[Online].Available:

https://oxfordmedicine.com/view/10.1093/med/9780190228484.001.0001/med-9780190228484-chapter-45

- [128] M. T. Foged *et al.*, 'Diagnostic added value of electrical source imaging in presurgical evaluation of patients with epilepsy: A prospective study', *Clinical Neurophysiology*, vol. 131, no. 1, pp. 324–329, Jan. 2020, doi: 10.1016/j.clinph.2019.07.031.
- [129] P. Mégevand and M. Seeck, 'Electric source imaging for presurgical epilepsy evaluation: current status and future prospects', *Expert Review of Medical Devices*, vol. 17, no. 5, pp. 405–412, May 2020, doi: 10.1080/17434440.2020.1748008.
- [130] Y.-H. Jun, T.-H. Eom, Y.-H. Kim, S.-Y. Chung, I.-G. Lee, and J.-M. Kim, 'Source localization of epileptiform discharges in childhood absence epilepsy using a distributed source model: a standardized, low-resolution, brain electromagnetic tomography (sLORETA) study', *Neurol Sci*, vol. 40, no. 5, pp. 993–1000, May 2019, doi: 10.1007/s10072-019-03751-4.
- [131] A. Coito *et al.*, 'Interictal epileptogenic zone localization in patients with focal epilepsy using electric source imaging and directed functional connectivity from lowdensity EEG', *Epilepsia Open*, vol. 4, no. 2, pp. 281–292, 2019, doi: https://doi.org/10.1002/epi4.12318.
- [132] W. Staljanssens *et al.*, 'EEG source connectivity to localize the seizure onset zone in patients with drug resistant epilepsy', *NeuroImage: Clinical*, vol. 16, pp. 689–698, Jan. 2017, doi: 10.1016/j.nicl.2017.09.011.
- [133] S. Vespa *et al.*, 'Ictal EEG source imaging and connectivity to localize the seizure onset zone in extratemporal lobe epilepsy', *Seizure*, vol. 78, pp. 18–30, May 2020, doi: 10.1016/j.seizure.2020.03.001.
- [134] E. Galaris, I. Gallos, I. Myatchin, L. Lagae, and C. Siettos, 'EEG source localization analysis in epileptic children during a visual working-memory task', *arXiv:2005.11186* [cs, math, q-bio], May 2020, Accessed: Feb. 14, 2021. [Online]. Available: http://arxiv.org/abs/2005.11186
- [135] J. Zhou, S. Liu, K. K. Ng, and J. Wang, 'Applications of Resting-State Functional Connectivity to Neurodegenerative Disease', *Neuroimaging Clinics*, vol. 27, no. 4, pp. 663–683, Nov. 2017, doi: 10.1016/j.nic.2017.06.007.
- [136] A. C. Tsolaki *et al.*, 'Brain source localization of MMN and P300 ERPs in mild cognitive impairment and Alzheimer's disease: a high-density EEG approach', *Neurobiol Aging*, vol. 55, pp. 190–201, Jul. 2017, doi: 10.1016/j.neurobiolaging.2017.03.025.
- [137] M. Hata *et al.*, 'Functional connectivity assessed by resting state EEG correlates with cognitive decline of Alzheimer's disease – An eLORETA study', *Clinical Neurophysiology*, vol. 127, no. 2, pp. 1269–1278, Feb. 2016, doi: 10.1016/j.clinph.2015.10.030.
- [138] F. La Foresta, F. C. Morabito, S. Marino, and S. Dattola, 'High-Density EEG Signal Processing Based on Active-Source Reconstruction for Brain Network Analysis in Alzheimer's Disease', *Electronics*, vol. 8, no. 9, Art. no. 9, Sep. 2019, doi: 10.3390/electronics8091031.
- [139] X. Kang *et al.*, 'Quantitative spatio-temporal characterization of epileptic spikes using high density EEG: Differences between NREM sleep and REM sleep', *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Feb. 2020, doi: 10.1038/s41598-020-58612-4.
- [140] A. Fernandez Guerrero and P. Achermann, 'Brain dynamics during the sleep onset transition: An EEG source localization study', *Neurobiology of Sleep and Circadian Rhythms*, vol. 6, pp. 24–34, Jan. 2019, doi: 10.1016/j.nbscr.2018.11.001.

- [141] A. Bersagliere, R. D. Pascual-Marqui, L. Tarokh, and P. Achermann, 'Mapping Slow Waves by EEG Topography and Source Localization: Effects of Sleep Deprivation', *Brain Topogr*, vol. 31, no. 2, pp. 257–269, Mar. 2018, doi: 10.1007/s10548-017-0595-6.
- [142] A. Brancaccio, D. Tabarelli, M. Bigica, and D. Baldauf, 'Cortical source localization of sleep-stage specific oscillatory activity', *Scientific Reports*, vol. 10, no. 1, Art. no. 1, Apr. 2020, doi: 10.1038/s41598-020-63933-5.
- [143] J. G. Klinzing *et al.*, 'Spindle activity phase-locked to sleep slow oscillations', *NeuroImage*, vol. 134, pp. 607–616, Jul. 2016, doi: 10.1016/j.neuroimage.2016.04.031.
- [144] M. Murphy, B. A. Riedner, R. Huber, M. Massimini, F. Ferrarelli, and G. Tononi, 'Source modeling sleep slow waves', *PNAS*, vol. 106, no. 5, pp. 1608–1613, Feb. 2009, doi: 10.1073/pnas.0807933106.
- [145] A. Castelnovo, B. A. Riedner, R. F. Smith, G. Tononi, M. Boly, and R. M. Benca, 'Scalp and Source Power Topography in Sleepwalking and Sleep Terrors: A High-Density EEG Study', *Sleep*, vol. 39, no. 10, pp. 1815–1825, Oct. 2016, doi: 10.5665/sleep.6162.
- [146] V. S. Handiru, A. P. Vinod, and C. Guan, 'Multi-direction hand movement classification using EEG-based source space analysis', in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Aug. 2016, pp. 4551–4554. doi: 10.1109/EMBC.2016.7591740.
- [147] M.-A. Li, Y.-F. Wang, S.-M. Jia, Y.-J. Sun, and J.-F. Yang, 'Decoding of motor imagery EEG based on brain source estimation', *Neurocomputing*, vol. 339, pp. 182– 193, Apr. 2019, doi: 10.1016/j.neucom.2019.02.006.
- [148] B. J. Edelman, B. Baxter, and B. He, 'EEG Source Imaging Enhances the Decoding of Complex Right-Hand Motor Imagery Tasks', *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 1, pp. 4–14, Jan. 2016, doi: 10.1109/TBME.2015.2467312.
- [149] Y. Hou, L. Zhou, S. Jia, and X. Lun, 'A novel approach of decoding EEG four-class motor imagery tasks via scout ESI and CNN', *J. Neural Eng.*, vol. 17, no. 1, p. 016048, Feb. 2020, doi: 10.1088/1741-2552/ab4af6.
- [150] J. M. Antelis, L. Montesano, A. Ramos-Murguialday, N. Birbaumer, and J. Minguez, 'Decoding Upper Limb Movement Attempt From EEG Measurements of the Contralesional Motor Cortex in Chronic Stroke Patients', *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 99–111, Jan. 2017, doi: 10.1109/TBME.2016.2541084.
- [151] A. Mottaz, M. Solcà, C. Magnin, T. Corbet, A. Schnider, and A. G. Guggisberg, 'Neurofeedback training of alpha-band coherence enhances motor performance', *Clinical Neurophysiology*, vol. 126, no. 9, pp. 1754–1760, Sep. 2015, doi: 10.1016/j.clinph.2014.11.023.
- [152] S. Dubovik *et al.*, 'EEG alpha band synchrony predicts cognitive and motor performance in patients with ischemic stroke', *Behav Neurol*, vol. 26, no. 3, pp. 187– 189, 2013, doi: 10.3233/BEN-2012-129007.
- [153] P. Nicolo, S. Rizk, C. Magnin, M. D. Pietro, A. Schnider, and A. G. Guggisberg, 'Coherent neural oscillations predict future motor and language improvement after stroke', *Brain*, vol. 138, no. Pt 10, pp. 3048–3060, Oct. 2015, doi: 10.1093/brain/awv200.
- [154] J. Friedrich and C. Beste, 'Paradoxical, causal effects of sensory gain modulation on motor inhibitory control – a tDCS, EEG-source localization study', *Scientific Reports*, vol. 8, no. 1, Art. no. 1, Nov. 2018, doi: 10.1038/s41598-018-35879-2.

- [155] X. Hong, Y. Wang, J. Sun, C. Li, and S. Tong, 'Segregating Top-Down Selective Attention from Response Inhibition in a Spatial Cueing Go/NoGo Task: An ERP and Source Localization Study', *Scientific Reports*, vol. 7, no. 1, Art. no. 1, Aug. 2017, doi: 10.1038/s41598-017-08807-z.
- [156] M. Sabeti, S. D. Katebi, K. Rastgar, and Z. Azimifar, 'A multi-resolution approach to localize neural sources of P300 event-related brain potential', *Computer Methods and Programs in Biomedicine*, vol. 133, pp. 155–168, Sep. 2016, doi: 10.1016/j.cmpb.2016.05.013.
- [157] M. J. Herrmann, A.-C. Ehlis, A. Muehlberger, and A. J. Fallgatter, 'Source Localization of Early Stages of Face Processing', *Brain Topogr*, vol. 18, no. 2, pp. 77– 85, Dec. 2005, doi: 10.1007/s10548-005-0277-7.
- [158] R. S. Andersen, A. U. Eliasen, N. Pedersen, M. R. Andersen, S. T. Hansen, and L. K. Hansen, 'EEG source imaging assists decoding in a face recognition task', in 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2017, pp. 939–943. doi: 10.1109/ICASSP.2017.7952294.
- [159] M. Calbi *et al.*, 'How context influences the interpretation of facial expressions: a source localization high-density EEG study on the "Kuleshov effect", *Scientific Reports*, vol. 9, no. 1, Art. no. 1, Feb. 2019, doi: 10.1038/s41598-018-37786-y.
- [160] R. Monteiro, M. Simões, J. Andrade, and M. Castelo Branco, 'Processing of Facial Expressions in Autism: a Systematic Review of EEG/ERP Evidence', *Rev J Autism Dev Disord*, vol. 4, no. 4, pp. 255–276, Dec. 2017, doi: 10.1007/s40489-017-0112-6.
- [161] R. K. Chikara and L.-W. Ko, 'Modulation of the Visual to Auditory Human Inhibitory Brain Network: An EEG Dipole Source Localization Study', *Brain Sciences*, vol. 9, no. 9, Art. no. 9, Sep. 2019, doi: 10.3390/brainsci9090216.
- [162] M. Fahimi Hnazaee, E. Khachatryan, and M. M. Van Hulle, 'Semantic Features Reveal Different Networks During Word Processing: An EEG Source Localization Study', *Front. Hum. Neurosci.*, vol. 12, 2018, doi: 10.3389/fnhum.2018.00503.
- [163] G. N. Dimitrakopoulos *et al.*, 'Task-Independent Mental Workload Classification Based Upon Common Multiband EEG Cortical Connectivity', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1940–1949, Nov. 2017, doi: 10.1109/TNSRE.2017.2701002.
- [164] I. Kakkos et al., 'Mental Workload Drives Different Reorganizations of Functional Cortical Connectivity Between 2D and 3D Simulated Flight Experiments', *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 27, no. 9, pp. 1704–1713, Sep. 2019, doi: 10.1109/TNSRE.2019.2930082.
- [165] K. Whittingstall, G. Stroink, L. Gates, J. Connolly, and A. Finley, 'Effects of dipole position, orientation and noise on the accuracy of EEG source localization', *BioMedical Engineering OnLine*, vol. 2, no. 1, p. 14, Jun. 2003, doi: 10.1186/1475-925X-2-14.
- [166] A. Delorme, T. Sejnowski, and S. Makeig, 'Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis', *NeuroImage*, vol. 34, no. 4, pp. 1443–1449, Feb. 2007, doi: 10.1016/j.neuroimage.2006.11.004.
- [167] C. M. Michel and D. Brunet, 'EEG Source Imaging: A Practical Review of the Analysis Steps', *Front. Neurol.*, vol. 10, 2019, doi: 10.3389/fneur.2019.00325.
- [168] X. Jiang, G.-B. Bian, and Z. Tian, 'Removal of Artifacts from EEG Signals: A Review', Sensors (Basel), vol. 19, no. 5, Feb. 2019, doi: 10.3390/s19050987.
- [169] Z. Akalin Acar and S. Makeig, 'Effects of Forward Model Errors on EEG Source Localization', *Brain Topogr*, vol. 26, no. 3, pp. 378–396, Jul. 2013, doi: 10.1007/s10548-012-0274-6.

- [170] S. Homölle and R. Oostenveld, 'Using a structured-light 3D scanner to improve EEG source modeling with more accurate electrode positions', *Journal of Neuroscience Methods*, vol. 326, p. 108378, Oct. 2019, doi: 10.1016/j.jneumeth.2019.108378.
- [171] C. C. Cline, C. Coogan, and B. He, 'EEG electrode digitization with commercial virtual reality hardware', *PLOS ONE*, vol. 13, no. 11, p. e0207516, Nov. 2018, doi: 10.1371/journal.pone.0207516.
- [172] W. H. Lee, Z. Liu, B. A. Mueller, K. Lim, and B. He, 'Influence of white matter anisotropic conductivity on EEG source localization: Comparison to fMRI in human primary visual cortex', *Clinical Neurophysiology*, vol. 120, no. 12, pp. 2071–2081, Dec. 2009, doi: 10.1016/j.clinph.2009.09.007.
- [173] V. Brodbeck *et al.*, 'Electroencephalographic source imaging: a prospective study of 152 operated epileptic patients', *Brain*, vol. 134, no. 10, pp. 2887–2897, Oct. 2011, doi: 10.1093/brain/awr243.
- [174] *MATLAB version 9.10.0.1613233 (R2021a)*. Natick, Massachusetts: The Mathworks, Inc., 2021.
- [175] A. Gramfort *et al.*, 'MEG and EEG data analysis with MNE-Python', *Frontiers in Neuroscience*, vol. 7, 2013, Accessed: Jul. 10, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2013.00267
- [176] C. R. Harris *et al.*, 'Array programming with NumPy', *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020, doi: 10.1038/s41586-020-2649-2.
- [177] P. Virtanen et al., 'SciPy 1.0: fundamental algorithms for scientific computing in Python', Nat Methods, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
- [178] J. D. Hunter, 'Matplotlib: A 2D Graphics Environment', *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, May 2007, doi: 10.1109/MCSE.2007.55.
- [179] M. L. Waskom, 'Seaborn: statistical data visualization', *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.
- [180] F. Pedregosa *et al.*, 'Scikit-learn: Machine Learning in Python', *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [181] 'Anaconda Software Distribution', *Anaconda Documentation*. Anaconda Inc., 2021. [Online]. Available: https://docs.anaconda.com/
- [182] NVIDIA, P. Vingelmann, and F. H. P. Fitzek, 'CUDA, release: 11.8'. 2021. [Online]. Available: https://developer.nvidia.com/cuda-toolkit
- [183] J. Bergstra and Y. Bengio, 'Random Search for Hyper-Parameter Optimization', *Journal of Machine Learning Research*, vol. 13, no. 10, pp. 281–305, 2012.
- [184] P. Liashchynskyi and P. Liashchynskyi, 'Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS', Dec. 12, 2019, arXiv: arXiv:1912.06059. doi: 10.48550/arXiv.1912.06059.
- [185] J. Snoek, H. Larochelle, and R. P. Adams, 'Practical Bayesian Optimization of Machine Learning Algorithms', in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2012. Accessed: Jan. 06, 2025. [Online]. Available: https://papers.nips.cc/paper\_files/paper/2012/hash/05311655a15b75fab86956663e181 9cd-Abstract.html
- [186] L. Bottou, 'Stochastic Gradient Descent Tricks', in *Neural Networks: Tricks of the Trade: Second Edition*, G. Montavon, G. B. Orr, and K.-R. Müller, Eds., Berlin, Heidelberg: Springer, 2012, pp. 421–436. doi: 10.1007/978-3-642-35289-8\_25.

- [187] K. He, X. Zhang, S. Ren, and J. Sun, 'Deep Residual Learning for Image Recognition', in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [188] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, 'Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)', Feb. 22, 2016, arXiv: arXiv:1511.07289. doi: 10.48550/arXiv.1511.07289.
- [189] Y. Nesterov, 'A method for unconstrained convex minimization problem with the rate of convergence o(1/k^2)', 1983. [Online]. Available: https://api.semanticscholar.org/CorpusID:202149403
- [190] A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 1st edition. Beijing; Boston: O'Reilly Media, 2017.
- [191] R. Kohavi, 'A study of cross-validation and bootstrap for accuracy estimation and model selection', in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*, in IJCAI'95. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., Aug. 1995, pp. 1137–1143.
- [192] F. Lotte *et al.*, 'A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update', *J. Neural Eng.*, vol. 15, no. 3, p. 031005, Apr. 2018, doi: 10.1088/1741-2552/aab2f2.
- [193] C. Vidaurre, C. Sannelli, K.-R. Müller, and B. Blankertz, 'Machine-learning-based coadaptive calibration for brain-computer interfaces', *Neural Comput*, vol. 23, no. 3, pp. 791–816, Mar. 2011, doi: 10.1162/NECO\_a\_00089.
- [194] S. Lemm, B. Blankertz, T. Dickhaus, and K.-R. Müller, 'Introduction to machine learning for brain imaging', *NeuroImage*, vol. 56, no. 2, pp. 387–399, May 2011, doi: 10.1016/j.neuroimage.2010.11.004.
- [195] P. Zanini, M. Congedo, C. Jutten, S. Said, and Y. Berthoumieu, 'Transfer Learning: A Riemannian Geometry Framework With Applications to Brain–Computer Interfaces', *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 5, pp. 1107–1116, May 2018, doi: 10.1109/TBME.2017.2742541.
- [196] W. Penfield and T. Rasmussen, 'The Cerebral Cortex of Man', Jason W. Brown Library, Jan. 1950, [Online]. Available: https://digitalcommons.rockefeller.edu/jasonbrown-library/5
- [197] H. H. Ehrsson, S. Geyer, and E. Naito, 'Imagery of voluntary movement of fingers, toes, and tongue activates corresponding body-part-specific motor representations', J *Neurophysiol*, vol. 90, no. 5, pp. 3304–3316, Nov. 2003, doi: 10.1152/jn.01113.2002.
- [198] A. M. Batula, J. A. Mark, Y. E. Kim, and H. Ayaz, 'Comparison of Brain Activation during Motor Imagery and Motor Movement Using fNIRS', *Comput Intell Neurosci*, vol. 2017, p. 5491296, 2017, doi: 10.1155/2017/5491296.
- [199] T. Hanakawa, I. Immisch, K. Toma, M. A. Dimyan, P. Van Gelderen, and M. Hallett, 'Functional Properties of Brain Areas Associated With Motor Execution and Imagery', *Journal of Neurophysiology*, vol. 89, no. 2, pp. 989–1002, Feb. 2003, doi: 10.1152/jn.00132.2002.
- [200] C. A. Porro *et al.*, 'Primary motor and sensory cortex activation during motor performance and motor imagery: a functional magnetic resonance imaging study', J *Neurosci*, vol. 16, no. 23, pp. 7688–7698, Dec. 1996, doi: 10.1523/JNEUROSCI.16-23-07688.1996.

- [201] N. Leeuwis, S. Yoon, and M. Alimardani, 'Functional Connectivity Analysis in Motor-Imagery Brain Computer Interfaces', *Front. Hum. Neurosci.*, vol. 15, Oct. 2021, doi: 10.3389/fnhum.2021.732946.
- [202] T. Hanakawa, M. A. Dimyan, and M. Hallett, 'Motor Planning, Imagery, and Execution in the Distributed Motor Network: A Time-Course Study with Functional MRI', *Cereb Cortex*, vol. 18, no. 12, pp. 2775–2788, Dec. 2008, doi: 10.1093/cercor/bhn036.
- [203] L. Zhavoronkova *et al.*, 'FMRI and EEG Reactions to Hand Motor Tasks in Patients with Mild Traumatic Brain Injury: Left-Hemispheric Sensitivity to Trauma', *Journal of Behavioral and Brain Science*, vol. 9, no. 6, Art. no. 6, May 2019, doi: 10.4236/jbbs.2019.96020.
- [204] V. Youssofzadeh *et al.*, 'Mapping and decoding cortical engagement during motor imagery, mental arithmetic, and silent word generation using MEG', *Hum Brain Mapp*, vol. 44, no. 8, pp. 3324–3342, Mar. 2023, doi: 10.1002/hbm.26284.
- [205] I. Giannopulu and H. Mizutani, 'Neural Kinesthetic Contribution to Motor Imagery of Body Parts: Tongue, Hands, and Feet', *Front. Hum. Neurosci.*, vol. 15, Jul. 2021, doi: 10.3389/fnhum.2021.602723.
- [206] A. G. He *et al.*, 'Modulation of neural connectivity during tongue movement and reading', *Hum Brain Mapp*, vol. 18, no. 3, pp. 222–232, Feb. 2003, doi: 10.1002/hbm.10097.

## Παραθέματα

## Κώδικας

Κώδικας 1 Αρχιτεκτονική του μοντέλου CNN1D\_Custom που παρουσιάστηκε στο Κεφάλαιο 5

```
1. class CNN1D Custom(nn.Module):
2.
       def __init__(self, n_timesteps, x_dim, y_dim, n_bands, n_outputs, dropout_rate=0.5,
batch norm=False):
            super(CNN1D_Custom, self).__init__()
3.
4.
            self.batch_norm = batch_norm
5.
             self.bn = nn.BatchNorm1d(n_timesteps, eps=1e-05, momentum=0.1) if batch_norm else
nn.Identity()
            self.ln1 = nn.LayerNorm(x_dim * y_dim * n_bands) if batch_norm else nn.Identity()
6.
7.
           self.ln2 = nn.LayerNorm([100,round(n_timesteps/2)]) if batch_norm else nn.Identity()
            self.ln3 = nn.LayerNorm([50,100]) if batch_norm else nn.Identity()
8.
            self.ln4 = nn.LayerNorm([50,50]) if batch_norm else nn.Identity()
9.
10.
            self.ap1 = nn.AvgPool1d(2)
            self.conv1 = nn.Conv1d(x_dim * y_dim * n_bands, 100, 5, padding='same')
11.
            self.conv2 = nn.Conv1d(100, 50, 5, padding='same')
12.
13.
            self.conv3 = nn.Conv1d(round(n_timesteps/2), 50, 5, padding='same')
            self.gap = nn.AdaptiveAvgPool1d(1) # Global Average Pooling
14.
15.
            self.dropout = nn.Dropout(dropout_rate)
            self.fc1 = nn.Linear(50, 100)
16.
17.
            self.fc2 = nn.Linear(100, n_outputs)
18.
19.
        def forward(self, x):
           x = self.ap1(x)
20.
           x = self.conv1(x)
21.
22.
           x = F.relu(x)
23.
           if self.batch_norm:
24.
                x = self.ln2(x)
           x = self.dropout(x)
25.
26.
            x = x.permute(0,2,1)
27.
           x = self.conv3(x)
28.
            x = F.relu(x)
           if self.batch_norm:
29
30.
               x = self.ln3(x)
31.
           x = self.dropout(x)
32.
           x = x.permute(0,2,1)
33.
            x = self.conv2(x)
34.
           x = F.relu(x)
35.
            if self.batch_norm:
              x = self.ln4(x)
36.
37.
           x = self.dropout(x)
38.
           x = self.gap(x).squeeze(-1)
39.
            x = self.fc1(x)
40
           x = F.relu(x)
           x = self.dropout(x)
41.
            x = self.fc2(x)
42.
43.
            return x
```

```
Κώδικας 2 Φόρτωση Δεδομένων με τη χρήση των εργαλείων MNE και Coleeg4 και η εφαρμογή source localization
 import numpy as np
 import os
 from importlib import reload
 from mne.datasets import fetch fsaverage
 from mne.beamformer import make lcmv, apply lcmv, apply lcmv epochs
 root folder= f'{os.path.expanduser(".")}'
 import numpy as np
 import os
 from importlib import reload
 # Importing coleeg functions
 sys.path.append(f'{root folder}')
 import coleeg pytorch as coleeg
 reload(sys.modules['coleeg pytorch'])
from coleeg pytorch import *
dataset = 'bcicomptIV2a'
start time = time.time()
unzip dataset(dataset=dataset)
print(f'Time = {timedelta(seconds=round(time.time() - start time))}')
 resample freq = None # resample frequency in Hz (Set to None to
 deactivate)
 resample freq = None # resample frequency in Hz (Set to None to
 deactivate)
 Bands = None
 # Bands = [(0.5, 8.0), (8.0, 13.0), (13.0, 40.0)] # Uncomment this line
 to divide signals into bands
 notch freqs = None # frequency (or list of frequencies) to which a notch
 filter is applied, example notch freqs = 50
 tmin = 0.0
 tmax = 2.0
 #_____
 _____
 # set sample frequency
 Sample_Freqs = { 'physionet':160, 'ttk':500, 'bcicomptIV2a':250,
 'chbmit':256}
 if resample freq is not None:
  Fs = resample_freq
 else:
  Fs = Sample Freqs[dataset]
 # Delete variables
 data_input,data_x, data_x2D, data_y, data_yc, train_x, train_x2D,
 train_y, train_yc, test_x, test_x2D, test_y, test_yc =
 0,0,0,0,0,0,0,0,0,0,0,0,0
 del data input, data x, data x2D, data y, data yc, train x, train x2D,
 train y, train yc, test x, test x2D, test y, test yc
```

```
# loading data
if dataset == 'bcicomptIV2a':
  import warnings
 Exclude=None
 warnings.filterwarnings("ignore", category=DeprecationWarning) #
disable annoying error messages from python
  data x, data y, data subject index =
get data bcicomptIV2a(resample freq=resample freq, Exclude=Exclude,
Bands=Bands, tmin=tmin, tmax=(tmax-1/Fs), Baseline=None)
else:
 print('Wrong dataset')
print(f'\nTime = {timedelta(seconds=round(time.time() - start time))}')
# normalizing data (mean => 0, standard deviation => 1)
normalize(data x)
# balancing data (for each subject, the number of each class is equal)
data x, data y, data subject index = balance(data x, data y,
data subject index)
mne.set log level("CRITICAL")
info = mne.io.read info(f'{dataset} info.fif')
fwd = mne.read forward solution(f'{dataset} ico4 fixed-fwd.fif')
src = fwd['src']
mne.set log level("CRITICAL")
info = mne.io.read info(f'{dataset} info.fif')
data x = np.squeeze(data x, axis=-1)
epochsAr = mne.EpochsArray(data x.transpose(0,2,1) , info)
subject = 'fsaverage'
fs dir = fetch fsaverage()
subjects dir = op.dirname(fs dir)
data cov = mne.compute covariance (epochsAr, tmin=0, tmax=0.5,
rank='full',
                                  method='empirical')
fwd = mne.read forward solution(f'{dataset} ico4 fixed-fwd.fif')
parc = 'aparc' # the parcellation to use, e.g., 'aparc' 'aparc.a2009s'
labels_parc = mne.read_labels_from_annot(
    subject, parc=parc, subjects_dir=subjects_dir)
new data x = []
labels parc = labels parc[:-1]
print('Applying source localization, using LCMV')
for i in tqdm(range(len(epochsAr)), desc="Processing epochs"):
    epochA = epochsAr[i]
    filters = make lcmv(epochA.info, fwd, data cov, reg=0.05,
                    pick ori='max-power',
                    weight norm='unit-noise-gain')
    [stc] = apply lcmv epochs(epochA, filters)
    del filters
    label ts = mne.extract label time course(
                    stc, labels parc, fwd['src'], mode='mean',
allow empty=True)
   new data x.append(label ts)
data = np.array(new data x).transpose(0,2,1)
data x = np.expand dims(data, axis=-1)
```